LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 69

Journal of
Experimental Psychology

Vol. 69, No. 1 January 1965

SEQUENTIAL PATTERNS AND MAXIMIZING[1]

CAMERON R. PETERSON[2] AND Z. J. ULEHLA[3]

University of Colorado

Most Ss in probability-learning experiments do not maximize, perhaps because they expect sequential patterns. The purpose of this experiment was to determine whether or not the elimination of the objective tenability of sequential dependencies would increase the proportion of maximizing responses. 21 Ss in the experimental condition controlled the random generation of events by the throw of a die so that sequential dependencies were objectively unreasonable. 21 control Ss were presented prearranged sequences, making it reasonable for S to anticipate sequential patterns. Results confirmed the experimental hypothesis; experimental conditions led to more maximizing responses than did control conditions at the .01 level of significance.

In typical probability-learning experiments, *S* can maximize the expected number of correct predictions by always predicting the most frequently occurring event; thus each prediction of the most frequent event can be termed a maximizing response. Most *S*s do not learn to maximize on every trial. Several hypotheses have been advanced as explanations of this suboptimal behavior. A major hypothesis is that *S*s respond not only to the unconditional event probabilities, but also respond to hypothesized sequential dependencies or patterns in the sequence of events (e.g., see Restle, 1961, pp. 107–113).

Three lines of evidence support the hypothesis that many *S*s do search for sequential patterns: (*a*) *S*s say that they are looking for patterns when protocols are obtained via the "thinking aloud" procedure (Feldman, 1962); (*b*) *S*s show recency effects (Jarvik, 1951) indicating that they respond to previous sequences of outcomes; and (*c*) *S*s are capable of finding patterns when they do exist in the sequence of outcomes (Anderson, 1960).

If hypothesized sequential patterns constitute one important reason for

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 41 of the Institute. This research was supported by a Research Grant (M-4977) from the National Institute of Mental Health. Grateful acknowledgment is made by the authors to Dorothy Boucher and Karene Will for running *S*s.

[2] Now at the University of Michigan.

[3] Now at the University of Denver.

lack of maximizing, then elimination of the objective basis for sequential patterns should increase maximizing. The purpose of this experiment was to compare maximizing in two conditions: (*a*) in the experimental condition *S* controlled the random generation of events so that he knew that no patterns were arranged by *E*; (*b*) in the control condition each sequence which emerged in the experimental condition was presented to a new *S* (here the sequences were determined in advance of presentation to *S*, making it objectively reasonable to anticipate sequential patterns). The experimental hypothesis was that the proportion of maximizing responses would be greater in the experimental condition.

METHOD

Experimental design.—The general experimental procedure was patterned after probability-learning experiments, except that the apparatus for the experimental group was designed to provide less basis for the existence of sequential dependencies than the apparatus for the control group. The essential difference was that the correct outcome ("black" or "white") on each trial was determined (*a*) in the experimental group by which side of a die rolled by *S* turned up and (*b*) in the control group by which color was on the back of a card turned up by *E*, the cards being stacked in a fixed sequence prior to the experiment.

Subjects and payment.—Forty-two University of Colorado students served as *S*s. Twenty-one *S*s participated in the experimental condition and 21 in the control condition.

The *S*s were paid 2 cents for each correct prediction and fined 2 cents for each incorrect prediction. Maximizing on every trial would yield an expected earning for the experiment of $2.00, more than any other behavior. For example, matching the event probabilities would yield an expected earning of $.67. The *S*s were obtained from the student employment office in order to enhance the utility of the earnings.

Apparatus.—The apparatus for the experimental conditions consisted of a die and a dice cup. Two sides of the die were covered with white tape and four sides were covered with black tape. The apparatus for the control condition consisted of a deck of 4 × 5½ in. opaque cards and a desk-calendar-type spindle which permitted the sequential ob-

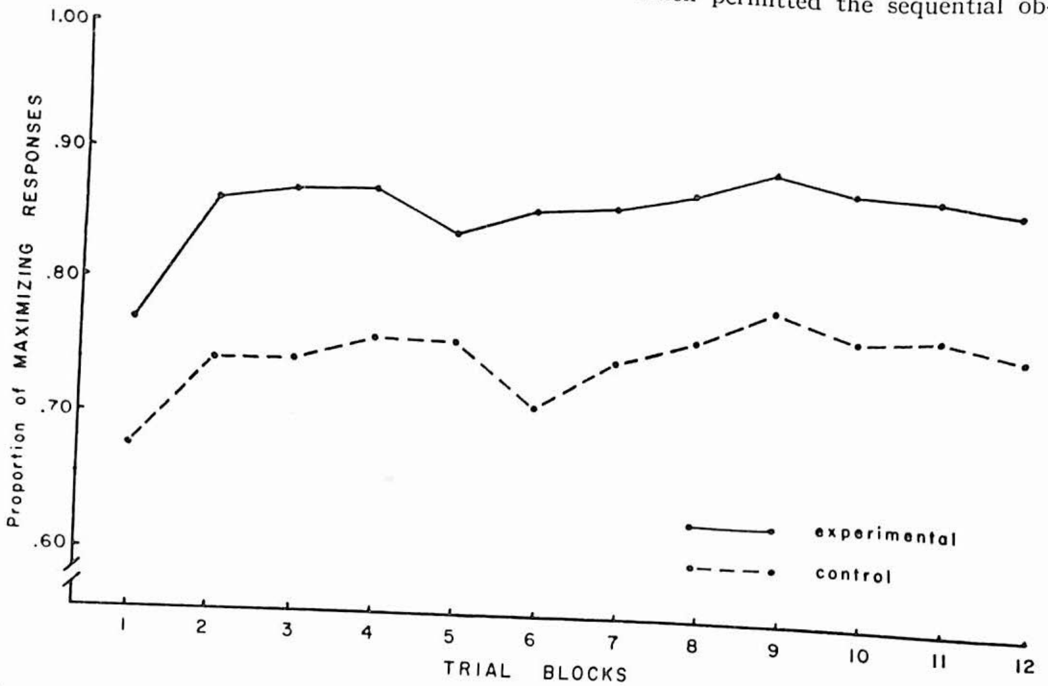


FIG. 1. Mean proportion of maximizing responses as a function of trial blocks for *S*s in the experimental group and for *S*s in the control group.

servation of the back of each card. A small square of either black tape or white was attached to the center of the back of each card. Four black and two white squares were displayed on the spindle as a constant reminder to S that there were approximately twice as many black as white squares in the deck.

Procedure.—An experimental trial consisted of requiring each S to: (*a*) predict which color would turn up upon a roll of the die, (*b*) roll the die, and (*c*) receive the 2-cent reward or penalty. A control trial consisted of requiring each S to: (*a*) predict which color would be on the back of the next card in the deck, (*b*) observe the correct color, and (*c*) receive the 2-cent reward or penalty.

The actual sequence of colors for each control S was determined as follows. One control S was paired at random with each experimental S. The sequence of colors in each control S's deck was prearranged to correspond with the sequence of colors actually rolled by the experimental S with whom he was paired. Each experimental and control S matched by sequence was considered as a matched pair where appropriate for the data analysis.

Results and Conclusion

Figure 1 presents the mean proportions of maximizing (black) predictions for the experimental and control groups during each of the 12 blocks of 25 trials each. The experimental group maintained approximately a .1 advantage throughout all blocks of trials. Since learning to maximize in both conditions seems to reach its peak after 200 trials, the last 100 trials were evaluated in terms of a t test, the results being $t\,(20) = 3.48$, $p < .01$.

Figure 2 presents the number of Ss who maximized completely and the

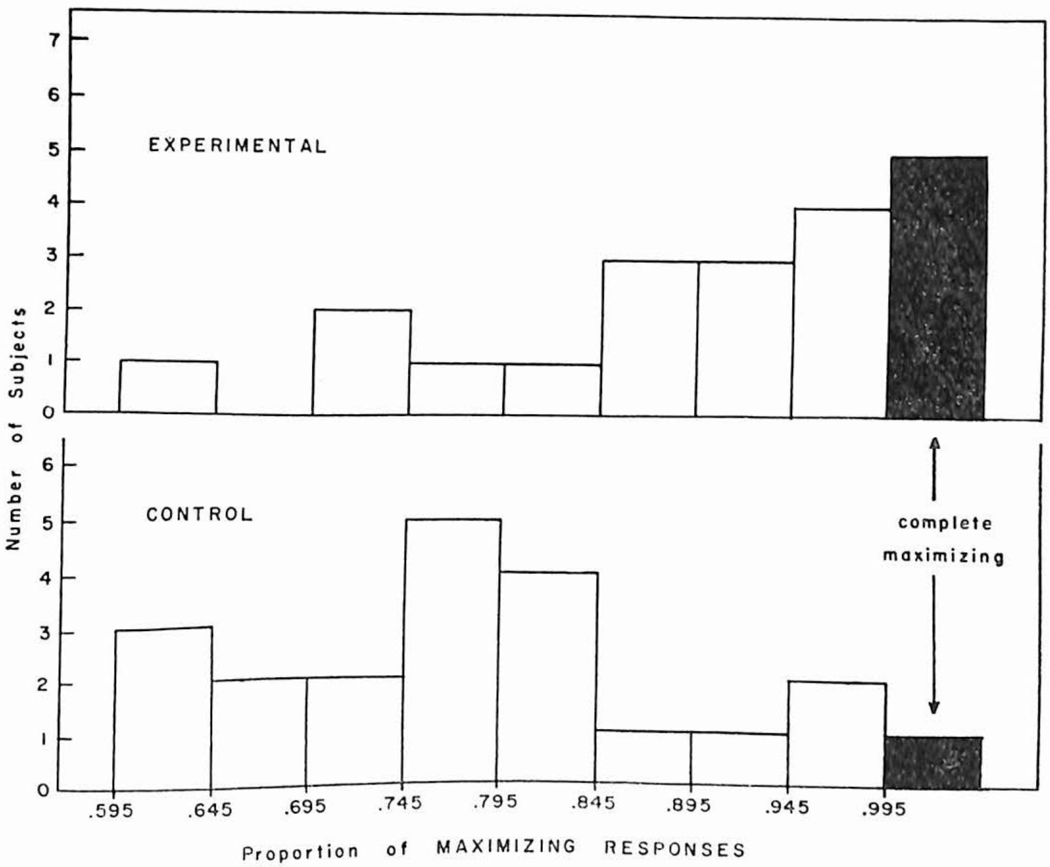


FIG. 2. Proportion of the last 100 trials on which individual experimental and control Ss maximized.

number of *S*s who fell into each .05 interval short of maximizing. More experimental than control *S*s exceed any arbitrary criterion of tendency to maximize between .65 and .99.

Clearly the experimental condition results in more maximizing than does the control condition. The experiment was designed to vary objective tenability of sequential dependencies. To the extent that this may be considered the essential difference between the experimental and control conditions, the experimental hypothesis is supported—elimination of an objective basis for sequential dependencies leads to more maximizing. This conclusion, in turn, supports the more general hypothesis that *S*s' anticipation of sequential dependencies is one of the factors leading to nonmaximizing responses in probability-learning experiments.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

FELDMAN, J. Computer simulation of cognitive processes. In H. Borko (Ed.), *Computer applications in the behavioral sciences.* Englewood Cliffs, N. J.: Prentice-Hall, 1962. Pp. 336–359.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

RESTLE, F. *Psychology of judgment and choice: A theoretical essay.* New York: Wiley, 1961.

(Received January 2, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 5–8

NUMBER AND TYPE OF AVAILABLE INSTANCES IN CONCEPT LEARNING [1]

VLADIMIR PISHKIN AND AARON WOLFGANG

Veterans Administration Hospital, Oklahoma City *University of Oklahoma School of Medicine*

This study explored the effects of availability to S of 2 types of stimulus instances, correctly and incorrectly sorted, on performance in a 4-choice concept-learning task. Errors to problem solution were reduced by the availability of 1 instance in each of the 4 response categories. Additional instances beyond 1 had no appreciable effect. Performance when only incorrectly sorted instances were exposed was unaffected by availability. There was no significant interaction between number and type of available instances.

The function of *S*'s memory in concept learning has been discussed recently by Hunt (1962, p. 143), who pointed out that, in Restle's (1960, 1961) model, concept learning may be assumed to occur without memory. The "no memory" model emphasizes hypothesis testing by *S* from a universe of potential hypotheses, where sampling of these is characterized by periodic selection and rejection of incorrect hypotheses. In this model the hypotheses simply continue to be replaced until *S* arrives at a correct one, although *S* may remember whether his working hypothesis has been correct for some of the preceding trials (Trabasso & Bower, 1964).

Somewhat contrary to the tenets of the "no memory" model is recent evidence that availability of instances from past trials facilitates concept performance (e.g., Cahill & Hovland, 1960).

Typically, stimuli to be categorized in concept learning are presented in succession and previous information must be retained in memory. Hovland and Weiss (1953) and Cahill and Hovland (1960) assessed the effect of memory for previously presented information by presenting the stimuli either *successively*, with no past information available, or *simultaneously*, with all past information exposed. The simultaneous procedure yielded more rapid learning than the successive presentation method. In a more recent experiment, Bourne, Goldstein, and Link (1964) investigated the number of available instances as a variable in concept learning. Groups which had up to five stimuli exposed tended to make fewer presolution errors than groups with less; there was a tendency for availability to produce a performance decrement with more than five stimuli exposed. The above studies demonstrate clearly that *S*s forget previously presented information and that *S* has at least two tasks in a concept problem; first, remembering information and secondly, on the basis of memory and of currently available stimuli, determining *E*'s classification rule.

Earlier concept-learning studies left available to *S* instances which had been categorized both correctly and incorrectly. The present study was designed to explore the functional relationships between concept learning and (*a*) *amount* (five levels) of available information from past trials

[1] Supported by Veterans Administration Medical Research Program (8,200 Funds). The valuable suggestions of Lyle E. Bourne, Jr. are gratefully acknowledged.

and (*b*) availability of three types of stimulus instances which were sorted by *S*s, viz., those sorted correctly, incorrectly, or a mixture of both correctly and incorrectly sorted instances.

METHOD

Subjects.—The *S*s were 90 male and female volunteer students from classes in elementary psychology at the University of Oklahoma. They were randomly divided into 15 treatment groups of 6 *S*s each.

Design.—A 3 × 5 factorial design was used. It included three conditions of type of available instances, viz., (PF) correctly sorted only, (NF) incorrectly sorted only, and (PNF) both types. Positive, negative, or combination of positive and negative feedback was given verbally by *E* in PF, NF, and PNF conditions, respectively. In addition, there were five conditions of number of available instances, i.e., zero, one, two, three, or four of the previously presented stimulus cards in each of the four categories were in *S*'s view on any trial. These conditions are referred to as n-0, n-1, n-2, n-3, and n-4, respectively. The *S*s in all conditions began by responding to a single stimulus. In availability conditions greater than n-0 this card remained in view while the next was presented. Once the specified level of availability was reached the earliest of all exposed stimuli was removed upon the presentation of each new one. Note that in Availability Cond. n-0 all levels of "type" of instances were treated identically, since no stimuli were available from previous trials.

Task and apparatus.—The *S*'s task was to categorize a series of geometric patterns in accordance with one relevant dimension. The relevant dimension for all *S*s was the form of the pattern. The test material was the Wisconsin Card Sorting Test, a pack of 64 geometric patterns, each containing one, two, three, or four identical figures (triangles, squares, crosses, and circles) in one of four colors (red, green, yellow, or blue). A gray wooden sorting tray, with four double compartments, was used. The upper half of the compartment was used for stimuli to which *S* made the wrong responses and the lower half for stimuli which *S* placed correctly. Each compartment was labeled A, B, C, or D.

Procedure.—At the start of each session *S* sat at a desk directly opposite *E*. The Wisconsin Card Sorting pack and the sorting tray were in front of *S*. The instructions for the PF condition follow:

> Listen carefully to the instructions. When I say "Begin," turn over the top card. Tell me where you think each of the cards should go, in Slot A, B, C, or D. I will tell you when you are right, and if you are right, lay the card face up in the slot you chose in the front row. The times you are not right, I will say nothing and you can lay the cards face down in the back row. Only one card (two, three, or four, depending on the condition) should be exposed at any one time in any one slot. Try to be right as often as possible. When you think that you have solved the problem stop and tell me what you think the answer is. (Face down was used in the condition where no past trials were exposed.)

The instructions for the NF condition were the same as above except that *S*s were told only when their responses were wrong and they were to lay the cards face up in the back row when wrong and face down in the front row when right; both front and back rows were exposed for PNF condition. The task ended when 64 trials had elapsed or when the *S* could correctly verbalize the solution; for example, crosses go in A, triangles in B, etc. There was no limit on the response-time interval.

RESULTS

The analysis of variance on the number of errors as a function of type and number of instances available was performed. Both the main effects of type, $F\ (2, 75) = 5.79$, $p < .01$, and number of available stimuli, $F\ (4, 75) = 2.85$, $p < .05$, were significant, but there was no Type × Availability interaction, $F\ (8, 75) = .81$, $p > .05$. Subsequent analyses with Duncan's test reveal, as does Fig. 1, that the greatest improvement in performance is from n-0 to n-1 level. In all comparisons, Cond. n-0 differed significantly from Cond. n-1, n-2, n-3, and n-4 ($df = 85$, $p < .05$). There were no significant differences between any other two levels of availability. Orthogonal polynomial analysis for the four components of trend was performed on the availability effect with

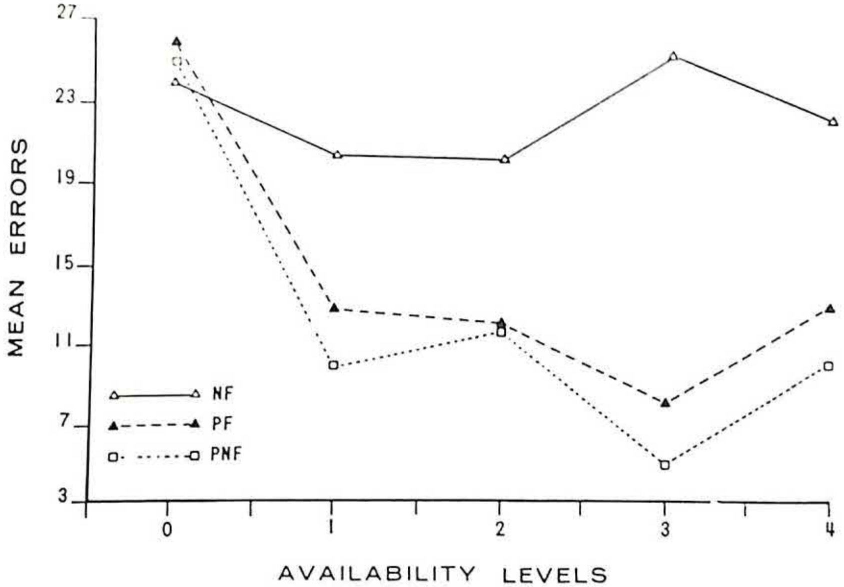


FIG. 1. Mean errors as a function of availability levels and three types of stimulus instances. (Each point represents N of 6.)

linear and quadratic components showing significance, as suggested by Fig. 1 ($F_{\text{lin}} = 5.67$, $F_{\text{quad}} = 4.61$, $p < .05$), which supports the conclusion that availability beyond n-1 is not facilitative.

The errors were most numerous in the NF group. Cross comparisons with Duncan's test between the three types of available instances showed that only the differences between NF and PF and between NF and PNF conditions were significant ($df = 87$, $p < .05$). Figure 1 shows that in the n-0 availability condition, where no previous information was available, type of instances was not an important factor; but, when information from past trials was provided, *S*s who were permitted to view only their wrong responses (NF) perform consistently worse across all availability levels than *S*s in PF and PNF groups. Thus, the overall significance of types of instances and availability levels is attributed to slower learning where (*a*) only incorrect instances were exposed and (*b*) where none of the previous stimuli were available.

DISCUSSION

The fact that PNF and PF groups with n-1 or more availability performed equally well and were superior to NF demonstrates that the type of stimulus instances available, rather than *E*'s feedback per se, facilitated solution. This conclusion emerges from the finding that, in the n-0 availability condition, *S*s did equally well, regardless of the type of feedback (Fig. 1).

The finding that availability of past information had a facilitative effect in concept learning was consistent with results of Hovland and Weiss (1953), Cahill and Hovland (1960), and Bourne, Goldstein, and Link (1964). Specifically, present evidence indicated that the greatest improvement in performance was from n-0 to Availability Level n-1 (up to four previous stimuli exposed). The relative simplicity of the problem used in the present study might serve to explain why *S*s failed to benefit from past available information beyond n-1.

Bourne et al. (1964) found that availability of previous stimuli was less facilitative in problems of lesser complexity. And since, in the present study an unlimited response interval was used, it may be ruled out that *S*s did not have sufficient time to assimilate information exposed at the higher availability levels.

With availability levels over *n*-1, PNF groups and PF groups were clearly superior in overall performance to NF groups. The result is likely due to *S*s having in view some correctly sorted instances, i.e., instances of what concept is, rather than incorrectly sorted instances which provide information what the concept is not. Formally, correctly sorted stimuli are positive instances and incorrectly sorted ones negative. It is well known that concept learning proceeds more rapidly on the basis of positive as compared to negative instances (Hovland & Weiss, 1953). That PF and PNF groups do equally well further testifies to the fact the *S*s in the PNF group gained very little information from having past errors available.

Facilitative effects of availability of "memory" information from past trials lend no support for the "no memory" assumption in concept-learning models. Present results stress at least limited utilization of previous instances, as was recently suggested by Trabasso and Bower (1964), who found that *S*s remembered information in their working hypothesis, since on solved problems, the recall of the relevant dimension was significant.

REFERENCES

Bourne, L. E., Jr., Goldstein, S., & Link, W. E. Concept learning as a function of availability of previously presented information. *J. exp. Psychol.*, 1964, **67**, 439–448.

Cahill, H. E., & Hovland, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

Hunt, E. B. *Concept learning.* New York: Wiley, 1962.

Restle, F. Note on the "hypothesis" theory of discrimination learning. *Psychol. Rep.*, 1960, **7**, 194.

Restle, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, **26**, 291–306.

Trabasso, T., & Bower, G. Memory in concept identification. *Psychon. Sci.*, 1964, **1**, 133–134.

(Early publication received June 22, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 9–24

RELATION BETWEEN STIMULUS INTENSITY AND OPERANT RESPONSE RATE AS A FUNCTION OF DISCRIMINATION TRAINING AND DRIVE [1]

JEFFREY A. GRAY [2]

Institute of Psychiatry, University of London

Noise intensity was varied while rats performed a bar-pressing response for food on a VI schedule. For 4 Ss (the Discrimination group) the noise was present only during periods of reinforcement; for 4 others (the Nondiscrimination group) noise was uncorrelated with reinforcement contingencies. Results indicate that response rate was a positive function of S^D intensity in the Discrimination group and essentially unrelated to noise intensity in the Nondiscrimination group. In the Discrimination group response rate during periods of S^Δ was a negative function of preceding S^D intensity. In a second experiment the effects of drive on these relationships were investigated. Although there were no significant interactions between drive and S^D intensity, the slopes of the curves relating response rate to noise intensity were systematically altered by variation in drive. Results are discussed in the light of Pavlovian theory and Perkins' (1953) generalization of inhibition account of stimulus-intensity dynamism.

In contrast to the large body of evidence gathered by Russian investigators of the classical conditioning situation that the intensity of the conditioned stimulus is an important determinant of the magnitude of the conditioned response (Gray, 1964; Razran, 1957), Western evidence on the same point remains scarce and, until recently, largely negative. A number of workers (Carter, 1941; Grant & Schneider, 1948, 1949; Kimmel, 1959; Kimmel, Hill, & Morrow, 1962; Passey & Herman, 1955; Wilcott, 1953) have reported failures to reproduce the phenomenon described in Pavlovian terms as "the law of strength," i.e. that, up to a limiting value of CS intensity, this parameter and response magnitude are positively related. Other workers have published results which are in agreement with the Russian findings, but have not found the law of strength to be a very clearly demonstrable phenomenon. Thus, Hovland (1937), a statistical analysis of whose results was performed by Hilgard and Marquis (1940), and Miller and Greene (1954) reported findings which were in the direction suggested by the Russian studies, but which failed to reach an acceptable level of statistical significance. Passey and Possenti (1956) found a difference, in the expected direction, in number and latency of alley-running responses for food reward in groups of rats trained at two fairly widely separated levels of illumination—4 and 128 ft-c—but found no differences between 8, 16, 32, and 64 ft-c. Similarly, Myers

[1] The work reported in this paper was supported by a grant from the Bethlem Royal and Maudsley Hospitals Research Fund. My thanks are due to P. L. Broadhurst for his advice and supervision, to H. J. Eysenck for his encouragement and for the provision of facilities in the Department of Psychology of the Institute of Psychiatry, and to H. Hurwitz for many helpful suggestions. P. Slater provided generous advice on statistical problems and N. Hemsley gave considerable assistance with the computations.

[2] Now at the Institute of Experimental Psychology, University of Oxford.

(1962), using a wheel-turning shock avoidance response with rats, found a greater number of responses to tones of 80 and 100 db. than to a tone of 60 db., but no differences either between the 80- and 100-db. tones or between 65- and 85-db. buzzers. Just as differences may be found between some stimuli, but not between others, so a law of strength may be found in the case of some individuals, but not in the case of others. Thus Zener (1949), using a classical salivary response with food reinforcement with five dogs, found a law of strength relating tone intensity to amount of salivation in the case of three of them, but not in the case of the other two. In five other studies (Barnes, 1956; Beck, 1963; Berlyne, 1950; Brown, 1942; Walker, 1960) which did succeed in demonstrating the operation of a law of strength, only two values of the stimulus were used, so that it is not possible to draw any conclusions from them about the nature of the function which relates conditioned stimulus intensity to response magnitude, other than that it is a positive one. Thus, the only study which, besides demonstrating the operation of a law of strength, also provides information at a number of intensities of the conditioned stimulus is that of Kessen (1953). This worker presented six light intensities, ranging from .16 to 2.26 log ft-c, as the CS for a wheel-turning response to avoid shock. He found that both number of CRs and response speed were positively related to CS intensity, and that the curves depicting this relationship rose to an asymptote at 1.90 log ft-c—the next to most intense light presented.

In view of this paucity of Western evidence for what Russian investigators confidently describe as a "law," and in view of the importance of this "law"—and, especially, of the exact nature of the function relating CS intensity to response magnitude—for certain aspects of Pavlovian personality theory (Gray, 1964), it seemed worth tackling the problem anew.

The great majority of previous Western studies of the effect of CS intensity on response magnitude have employed aversive stimulation (Barnes, 1956; Beck, 1963; Kamin & Schaub, 1963; Kessen, 1953; Myers, 1962; Walker, 1960; and others). This is an unfortunate choice, for high intensities of any kind of stimulation may be noxious and therefore elicit fear responses; thus, an apparent effect of CS intensity on the magnitude of an aversive response may be due to the unconditioned fear-inducing properties of the CS rather than to its role as a conditioned signal. For this reason, the present investigation employs an appetitive response—bar pressing for a food reward by the hungry rat. White noise, at five intensities ranging from 70 to 100 db., served as S^D.

Experiment I

In at least two theoretical contexts we would expect that only discriminative stimuli would affect response rate. Pavlovian theory (Gray, 1964) and the Russian empirical data (Razran, 1957) indicate that the law of strength only appears when "excitation is concentrated"—which means, in effect, that it is only a *conditioned* stimulus which is able to affect response magnitude. The same problem has been raised in Western work on "stimulus intensity dynamism," Hull's (1949) synonym for Pavlov's "law of strength." Perkins (1953) has argued that stimulus intensity dynamism is due to generalization of inhibition from an unreinforced zero-intensity stimulus, and therefore only

appears in experimental situations in which the reinforcement contingencies deliberately or unwittingly set up by *E* determine the formation of a discrimination between a positive stimulus and the absence of that stimulus.

Perkins' generalization of inhibition hypothesis supposes a failure of discrimination between a low-intensity S^D and an S^Δ constituted by the absence of the S^D. Clearly, such a failure of discrimination should work both ways: if the negative stimulus is presented after the positive one, it should elicit a *higher* rate of response, the *lower* the intensity of the positive stimulus—a kind of law of strength in reverse.

Experiment I is concerned, then, with the following problems: (*a*) Is it possible to demonstrate a clear-cut law of strength in the situation employed? (*b*) Is this effect limited to a discriminative stimulus or does it also occur with nonsignal stimuli? (*c*) Is there an effect of S^D intensity on response rate in a subsequent period of S^Δ?

Method

Subjects

The *S*s were eight experimentally naive adult male albino rats belonging to the Maudsley Nonreactive strain (designated MNR by Jay, 1963) and between 116 and 160 days old at the beginning of the experiment. The Discrimination group consisted of four *S*s (r 27, 37, 18, and 12) and the Nondiscrimination group of another four (r 9, 10, 11, and 20).

Apparatus

The apparatus consisted of: (*a*) a modified Skinner box; (*b*) a system of relays and a film-strip timer which enabled *E* to program either fixed interval (FI) or variable interval (VI) reinforcement of the animal's lever-pressing responses, together with counters which recorded the number of responses and the number of reinforcements; and (*c*) a sound-generation system which delivered white noise of variable intensity to four 8-in. loudspeakers mounted on top of an external box in which the Skinner box was placed. The Skinner box, 8 × 13 × 8 in., contained a single lever and a food tray attached to the rear wall. A force of 7–8 gm. was sufficient to depress the lever. Reinforcement consisted of pellets weighing approximately .05 gm., and containing 50% rat cake meal (MRC diet 41B) and 50% sucrose. The pellets were delivered to the food tray by a post office uniselector. The Skinner box and the uniselector were placed in an external box, 8 × 16 × 16 in., made of ½-in. thick softwood. The four loudspeakers were rigidly mounted on top of the external box. No attempt was made to soundproof the apparatus, but the two boxes in which the experimental animal was placed were kept on their own in a relatively quiet room (ambient noise level did not usually exceed 60 db.).

Noise intensities were calibrated by means of a Dawe sound-level meter, Type 1400D, in db. re 0.0002 dynes/cm². For this purpose, the microphone of the sound-level meter was placed on the floor of the Skinner box, midway between the lever and the food tray, roughly where the animal's head would normally be while it was responding on the lever. Readings obtained with the sound-level meter were correlated with the readings obtained on a voltmeter across the amplifier output, both while this was being delivered to the loudspeakers and while it was being delivered to a dummy load. In this way, by using the voltmeter readings, it was possible for *E*, during the course of an experiment, to select the required intensity for the next period of noise (see Procedure) while no noise was being delivered to the Skinner box, and also to monitor the intensity of the noise actually delivered to the Skinner box. Weekly calibration checks showed that the error attached to this method of selecting the required noise intensity was usually within the limits of ±1 db.

Procedure

Throughout the experiment, testing was conducted from Monday to Friday only; no testing occurred at weekends, though deprivation schedules were maintained. The *S*s were put on a 22-hr. deprivation schedule, and were fed daily between 3 and 5 P.M. Testing was always carried out in the morning, usually between 9.30 A.M. and 12 noon for the Discrimination group, and between 12 and 1 P.M. for the Nondiscrimination group.

Preliminary training.—After *S*s' weights had stabilized on the deprivation schedule,

they were introduced to the Skinner box for habituation and measurement of operant level. Each *S* was placed in the apparatus on 5 consecutive days and exposed to the same stimulation by noise to which it was to be exposed during the main part of the experiment (see below) but without reinforcement. (Owing to a later change in the composition of the groups, Rat r 12 was treated as a member of the Nondiscrimination group during habituation.) The number of times *S* pressed the lever during each period of stimulation by noise or by silence was recorded.

There followed 3 wk. of training on the bar-pressing response for food reinforcement. By the end of the first of these 3 wk. all *S*s were working successfully on FI 7.5 sec.; for the remaining 2 wk. they were on VI 30 sec., each daily session lasting 12.5 min. During this part of the experiment no noise was presented to *S*s. Response and reinforcement scores were recorded at the end of every 2.5 min.

Main experiment.—After completion of the 4 wk. preliminary training, the two groups were placed on different schedules. For the Discrimination group periods of noise alternated regularly with periods of silence, all periods lasting 2 min., 45 sec.; during periods of noise (S^D), reinforcement was available on VI 30 sec.; during silence (S^Δ), no reinforcement was delivered. A daily session lasted 27.5 min. and consisted of five periods of VI and five of extinction. Sessions always began with VI and ended with extinction. The Nondiscrimination group continued on VI 30 sec., with no interspersed extinction components. A session for this group consisted of five periods, each lasting 2 min., 45 sec. and each accompanied by noise; thus the session length was 13 min., 45 sec. For both groups five noise intensities were used: 70, 80, 85, 90, and 100 db. In a single daily session, each of these intensities was presented once in a random order determined by a 5 × 5 Latin square, of which days of the week formed the columns; periods within a session, the rows; and noise intensities, the letters. To ensure the maximum randomization, a separate Latin square was selected at random for each *S* each week. For the Discrimination group, *E* was able to adjust the volume control during a period of silence; in order to do the same for the Nondiscrimination group, it was necessary to turn off the noise 7.5 sec. before the end of each noise period and adjust the volume control in this interval of silence. Any reinforcements programed during this interval were delivered. Scores for the relevant noise period were taken when the subsequent intensity was turned on; a period of noise for the Nondiscrimination group thus, in fact, consisted of 2 min., 37.5 sec. noise followed by 7.5 sec. silence.

During each period of VI, five reinforcements were programed; maximum possible reinforcement during a single session was thus 25 pellets, and it was rare for any *S* not to obtain all of these. All *S*s were tested for at least 4 wk. in the ways described; two members of the Discrimination group (r 27 and 37) were tested for a fifth week as well.

Three *S*s from the Nondiscrimination group (r 9, 10, and 11) were tested for a further 2 wk. with a changed procedure. Session length was doubled, along with maximum possible reinforcements, and periods of noise were alternated with periods of silence, all periods lasting 2 min., 45 sec. and reinforcement being available throughout. Noise intensity was randomly allocated to period as before.

Throughout all the various treatments, *E* recorded the number of responses and of reinforcements at the end of each of the periods into which sessions were divided.

Results

Operant Level and Discrimination Training

Analysis of variance of the session totals for noise and silence obtained from r 27, 37, and 18 during the week of habituation showed that the presence of noise in the experimental space had a dynamogenic effect on bar-pressing behavior, $F\ (1, 8) = 8.36$, $p < .05$: the mean number of responses per session per *S* in noise was 7.6, in silence 3.9. Analysis of variance of the scores obtained during periods of noise by all eight *S*s showed that noise intensity had no effect on operant level in either group. No source of variation was significantly different from chance in this analysis except Animals. A *t* test showed that r 20 had a significantly higher score than any other *S*; the other *S*s did not differ from one another.

There was no difference between the rates of response achieved by the

TABLE 1

Response Rates per Minute in Noise and Silence in the Discrimination Group in Exp. I and II

	Exp. I		Exp. II Low Drive		Exp. II High Drive		% Increase from Low to High Drive	
	Noise	Silence	Noise	Silence	Noise	Silence	Noise	Silence
r 27	37.61	10.82	35.80	5.13	48.85	12.21	36	138
r 37	20.31	5.66	17.06	2.65	22.65	4.39	33	66
r 18	25.22	13.13	54.24	18.63	59.20	24.41	9	31
r 12	23.98	9.34	29.61	6.60	36.80	13.49	24	104

two groups by the end of the training period, nor were rates affected differently in the two groups by the transition to the final phase of the experiment.

In the Discrimination group, all four *S*s learnt the discrimination between noise and silence. In each case, response rate during noise was higher than response rate during silence at levels of significance well beyond $p = .001$ (Table 1). Furthermore, in the case of three of the four *S*s, the discrimination was clearly established by the end of the first week: although the Weeks × Noise term in the analyses of variance was significant in the case of r 37, r 18, and r 12, *t* tests showed that rate in noise was higher than rate in silence for all weeks of testing except in the case of r 18; r 18 did not show a significant difference between rate in noise and rate in silence until Week 2. The curves for response rate in noise and in silence are shown, for each *S*, in Fig. 1. If it is clear that all *S*s successfully discriminate noise from silence, the question still arises whether they succeeded in discriminating the lowest intensity noise—70 db.—from silence. To answer this question, analysis of variance was used to compare the scores obtained during the 70-db. periods with the mean daily silence-period scores (i.e., total session scores during silence divided by five, since there were five periods of silence per session). The results of the analyses showed that the rate of responding during 70-db. noise was significantly higher, at beyond the 0.1% level, than the rate during silence for each *S*. In the case of r 37 and r 18, there was a significant interaction between Weeks and Noise (i.e. 70 db. noise); *t* tests showed that rate during 70 db. was

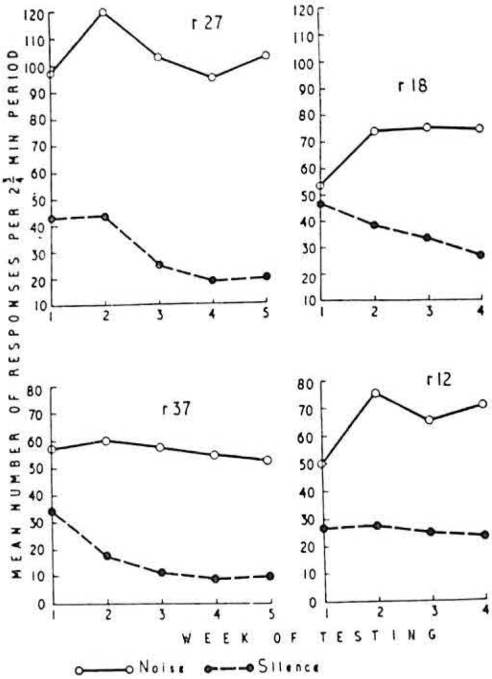


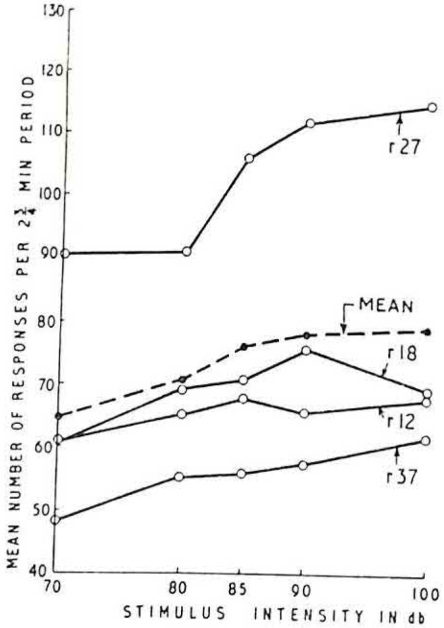
Fig. 1. Response rates in noise (S^D) and silence (S^Δ) for each member of the Discrimination group in Exp. I.

higher than rate during silence for all 4 wk. in the case of r 37, but not in the case of r 18, which did not succeed in discriminating 70 db. from silence until the second week of testing.

Effects of Stimulus Intensity on Concurrent Reinforced Responding

Figure 2 shows response rate as a function of noise intensity in the two groups; Table 2 presents, for each *S* separately and for the two groups, those results of the analyses of variance of the scores recorded during noise periods which concern the Intensity variable. As well as the significance of the main Intensity effects, Table 2 also presents the results of tests of the significance of linear regression of response rate on noise intensity and, where appropriate, regression equations. It will be seen that, in the Discrimination group, response rate is in all cases a significant positive function of S^D intensity; linear regression is significant in the case of each *S*; r 27 and the group as a whole show a curvilinear component as well. In the Nondiscrimination group, on the other hand, response rate remains unaffected by noise intensity in two *S*s, and in the group considered as a whole. There is, however, a weakly significant effect of noise intensity in the case of two *S*s (r 9 and 10), but in neither case does this effect resemble the effect of noise intensity in the Discrimination group: r 9 tends to respond more the lower the intensity to which it is exposed, while r 10 presents a curvilinear relationship between intensity and response rate, rising to a peak at 85 db. and then falling again.

The effects of noise intensity in the Discrimination group remained the same throughout the experiment, as indicated by nonsignificant interactions between Weeks and Intensities. In the case of two members of the Nondiscrimination group, however, there were significant interactions of



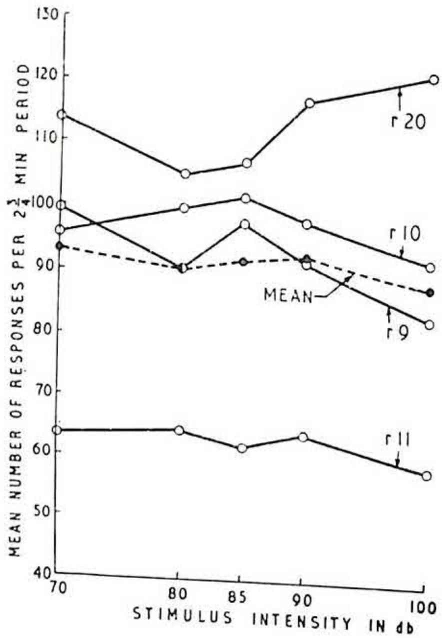


FIG. 2. Response rate during noise as a function of noise intensity in Exp. I: left-hand graph, the Discrimination group; right-hand graph, the Nondiscrimination group.

TABLE 2

Effects of Noise Intensity on Response Rate during Presentation of Noise: Significance Levels and Linear Regression Equations in Exp. I

S	Main Effect F	Linear Regression F	Departures from Linearity F	Regression Equation
Discrimination Group				
r 27	16.10***	56.02***	3.43*	$y = 20.8 + 0.972x$
r 37	11.58***	44.05***	0.76	$y = 18.8 + 0.437x$
r 18	3.21*	5.69*	2.38	$y = 40.5 + 0.339x$
r 12	3.65*	11.42**	1.06	$y = 45.3 + 0.243x$
Group	23.04***	81.48***	3.56*	$y = 29.9 + 0.519x$
Nondiscrimination Group				
r 9	3.14*	8.41**	1.39	$y = 132.3 - 0.467x$
r 10	3.40*	1.39	4.07*	—
r 11	1.48	—	—	—
r 20	0.79	—	—	—
Group	0.66	—	—	—

Note.—$df = 4$ for the main effect, 1 for linear regression, and 3 for departures from linearity; the error variances have $df = 48$ in the case of the individual Ss ($df = 60$ for r 27 and r 37) and $df = 192$ in the case of the group analyses. In the regression equations, y = mean number of responses in a single period of noise lasting 2 min., 45 sec. and x = noise intensity in db.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

this kind: for r 10, $F\ (12, 48) = 2.78$, $p < .01$; for r 11, $F = 3.60$, $p < .001$. It was difficult to discern any systematic trend in the data producing these interactions; there was, however, a suggestion of a predominantly negative relationship between stimulus intensity and response rate in the first week or two which then disappeared. There was only one other significant interaction involving Intensity in any of the analyses. This was an interaction between Animals and Intensity in the analysis of the Discrimination group, $F\ (12, 192) = 4.22$, $p < .001$. When t tests were carried out, it appeared that the most important differences going to make up this interaction were the much steeper rise between 80 and 85 db. shown by r 27 than by any other S and the significant ($p < .05$) fall between 90 and 100 db. shown by r 18 (Fig. 2).

Analysis of variance of the daily session totals for noise and silence recorded for the three Ss (r 9, 10, and 11) from the Nondiscrimination group exposed for an extra 2 wk. of testing to alternating noise and silence with reinforcement available throughout (see Procedure) demonstrated that, in each case, rate during noise was significantly higher than rate during silence (Table 3). When the scores obtained from these Ss during the periods of noise were submitted to analysis of variance, it was found that noise intensity did not affect response rate in any S, nor were there any significant interactions between Intensity and Week of Testing. When the results of the three Ss were analyzed together, however, a significant effect of Intensity appeared, $F\ (4, 72) = 4.16$, $p < .01$; there were no significant interactions between Intensity and either Weeks or Animals. Figure 3 shows both the

TABLE 3

DIFFERENCES BETWEEN RESPONSE RATES DURING NOISE AND DURING SILENCE IN THREE Ss FROM THE NONDISCRIMINATION GROUP EXPOSED TO ALTERNATING NOISE AND SILENCE

Animal	Rate per Minute during Silence	Rate per Minute during Noise	F
r 9	32.14	39.79	20.80*
r 10	44.21	47.10	17.48*
r 11	31.53	35.46	64.66**

Note.—df for F values = 1/4.
* $p < .05$.
** $p < .01$.

significant Intensity effect for the group and the nonsignificant effects for each S. The curvilinearity of the Intensity effect for the group, obvious from the figure, was confirmed by regression analysis, which showed that only the component of variation due to departures from linearity was significant, $F(3, 72) = 5.52$, $p < .01$.

Effects of Stimulus Intensity on Responding after the Termination of the Stimulus

Figure 4 presents the relationship between noise intensity and response rate during the subsequent period of

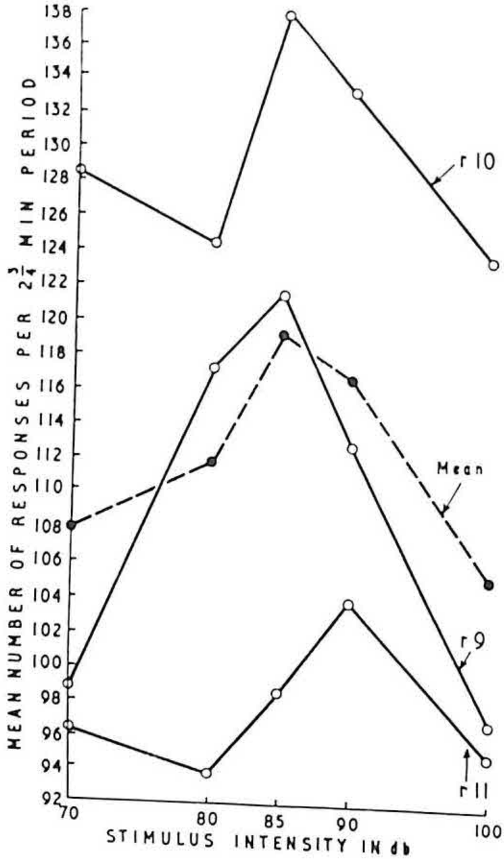


FIG. 3. Response rate during noise as a function of noise intensity in three members of the Nondiscrimination group exposed to alternating noise and silence with reinforcement available in both.

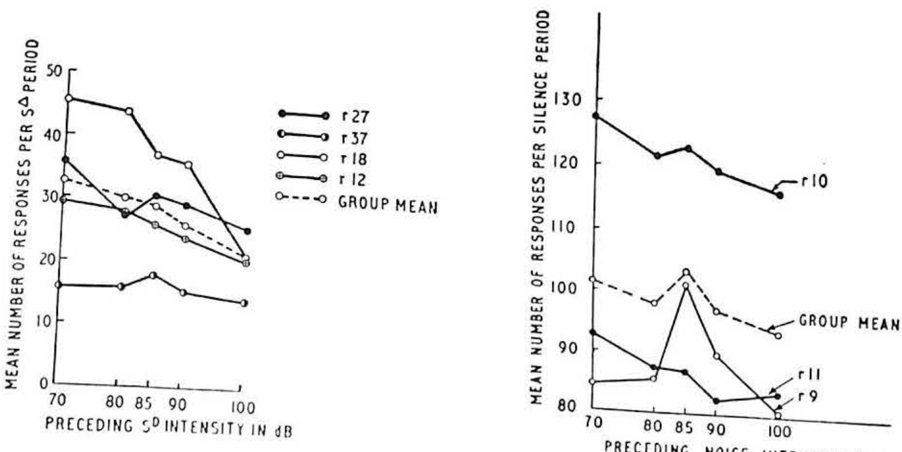


FIG. 4. Response rate during silence as a function of preceding noise intensity in Exp. I: left-hand graph, the Discrimination group; right-hand graph, the Nondiscrimination group.

TABLE 4

EFFECTS OF NOISE INTENSITY ON RESPONSE RATE DURING SUBSEQUENT PERIODS OF SILENCE: SIGNIFICANCE LEVELS AND LINEAR REGRESSION EQUATIONS IN EXP. I

S	Main Effect *F*	Linear Regression *F*	Departures from Linearity *F*	Regression Equation
Discrimination Group				
r 27	2.46	6.50*	1.11	$y = 53.1 - 0.275x$
r 37	0.56	0.52	0.57	—
r 18	9.78***	35.73***	1.13	$y = 108.9 - 0.857x$
r 12	2.91*	11.14**	0.17	$y = 49.6 - 0.281x$
Group	10.92***	33.56***	3.36*	$y = 52.6 - 0.292x$
Nondiscrimination Group				
r 9	0.80	0.03	1.05	—
r 10	0.64	2.33	0.07	—
r 11	1.41	4.56*	0.36	$y = 114.1 - 0.323x$
Group	1.20	2.40	0.80	—

Note.—$df = 4$ for the main effect, 1 for linear regression, and 3 for departures from linearity; the error variances have $df = 60$ in the case of r 27 and r 37, 48 for r 18 and r 12, 24 for r 9, r 10, and r 11, 192 for the Discrimination group, and 72 for the Nondiscrimination group. In the regression equations, y = mean number of responses in a single period of silence lasting 2 min., 45 sec. and x = preceding noise intensity in db.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

silence for the Discrimination group and for the Nondiscrimination group during exposure to alternating noise and silence. Table 4 shows the significance levels of the Intensity effects illustrated by the curves in Fig. 4. It will be seen that, in the Discrimination group, every *S* except r 37 shows a significant linear negative relationship between S^D Intensity and subsequent S^Δ response rate. The same pattern also appeared in the analysis of the Discrimination group as a whole, although, as well as the linear component of variation, there were significant departures from linearity, apparently due to a steeper decline in response rate between 85 and 100 db. than between 70 and 85 db. The group analysis also revealed a significant Animals × Intensity interaction, $F(12, 192) = 3.04$, $p < .001$; this seems to be due to the particularly marked effect of S^D intensity on S^Δ response rate displayed by r 18 and the lack of any effect in r 37. The curves for the Nondiscrimination group are not dissimilar to those for the Discrimination group; however, only in the case of r 11 was the effect of noise intensity weakly significant when linear regression was taken into account. Since there were no significant effects of noise intensity, either alone or in interaction, in the group analysis, the most likely conclusion is that this variable, when it has no signal properties, has no real effect on response rate in silence.

EXPERIMENT II

Pavlovian theory (Gray, 1964) predicts that the function relating stimulus intensity to conditioned response magnitude is itself a function of certain other variables. Among the most important of these are drive

level and individual differences along a dimension of personality termed the "strength of the nervous system." The positive relationship between stimulus intensity and response magnitude is supposed to hold only up to a limiting value of stimulus intensity; beyond this point, further increases in stimulus intensity are reported to cause a decrement (or, at least, to cease causing an increase) in response magnitude. This limiting value of stimulus intensity is known as the "threshold of transmarginal inhibition." Pavlovian theory holds that this threshold will be lower, the higher the drive level. Experiment II reports a test of this hypothesis: curves relating S^D intensity to operant response rate are obtained from the same *S*s under two levels of drive which differ both from one another and from the drive level obtaining in Exp. I.

Method

Subjects

The *S*s were the four members of the Discrimination group and r 20 from the Nondiscrimination group used in Exp. I. All *S*s were placed on the new feeding schedule described in the procedure below. In the case of r 27 and 37, 1 wk. with no testing of any kind, but with the new feeding schedule in operation, intervened between the two experiments; for the other three *S*s there was only the usual weekend break, during which the new schedule was commenced.

Procedure

The apparatus has been described in Exp. I. The reinforcement contingencies remained as before.

All *S*s were placed on a new deprivation schedule. On this schedule they were fed for 1 hr. a day only, either between 9 and 10 A.M. or between 12 and 1 P.M. On days on which *S* was fed between 9 and 10 A.M., it was tested between 12 and 1 P.M.; these constituted the "low-drive" days. On days on which *S* was fed between 12 and 1 P.M., it was tested between 9.30 and 10.30 A.M.; these constituted the "high-drive" days. Thus, if we adopt, with some slight changes, Eisman's (1956) symbolism, in which h = hours since last feeding and h_t = hours per 24 hr. during which the animal is allowed ad-lib. access to food, then:

In Exp. I, $h = 16.5 - 19.5$, $h_t = 2$.

In the high-drive condition of Exp. II, h = app. 24, $h_t = 1$.

In the low-drive condition of Exp. II, h = app. 2, $h_t = 1$.

For each *S*, low- and high-drive days alternated regularly; the first day of the first week of testing was a low-drive one for r 27 and r 37 and a high-drive one for r 18, r 12, and r 20. Apart from the change in drive conditions, the procedure was the same as in Exp. I. Random presentation of noise intensities during the periods into which daily sessions were divided was again achieved by preparing a 5 × 5 Latin square for each animal. Each Latin square was now, however, repeated for 2 consecutive wk. In this way, each column of the Latin square was presented both on a low-drive day and on a high-drive day. For example, if H = high drive and L = low drive, and the numbers 1–5 stand both for the five columns of the Latin square and the 5 days of the week on which testing was conducted, then a typical 2-wk. testing sequence would read:

```
H L H L H L H L H L H L H L
1 2 3 4 5       1 2 3 4 5
```

Over the weekend, although no testing occurred, the alternation of feeding times continued. The experiment was continued for 4 wk., allowing one replication of the testing sequence described above; a different Latin square was used for each replication.

Results

Effect of Drive Level on Response Rates

The method adopted for varying drive level was successful in the case of all *S*s. In each case, response rates in both noise and silence were higher on high-drive days than on low-drive days at levels of significance beyond $p = .001$ (by analysis of variance). Mean response rates under the two conditions are shown for the Discrimination group in Table 1. As

might be expected on the assumption that time since last feeding (h) is the main determinant of drive level, the response rates recorded in Exp. I fall between the values recorded in the two drive conditions of Exp. II in the case of r 27 and 37 for S^D responding and r 27 and 12 for S^Δ responding. Rat r 20 also displayed this ordering of response rates (35.93 responses/min in the low-drive condition, 41.18 in Exp. I, and 48.48 in the high-drive condition). However, both the low- and the high-drive conditions generated higher response rates than those recorded in Exp. I during both S^D and S^Δ in the case of r 18 and during S^D only in the case of r 12. This overall increase in response rate could be due to continued experience of the experimental situation, to the shorter time allowed for adaptation to the new feeding schedule, or to a particularly marked effect in these *S*s of the reduction in number of hours feeding per 24 hr. (h_t; Eisman, 1956). In the case of r 37, due, no doubt, to continued nonreinforcement, S^Δ response rate was lower in both drive conditions than in Exp. I. All these differences between response rates in the two experiments were fully significant, as determined by analysis of variance.

Effects of Noise Intensity and Drive in the Discrimination Group

Figure 5 presents the curves relating both S^D and S^Δ response rates to S^D intensity for each *S* in the Discrimination group. Table 5 shows the results of the relevant significance tests from the analyses of variance together with the linear regression equations. It can be seen that the results of this experiment confirm both the positive effect of S^D intensity on

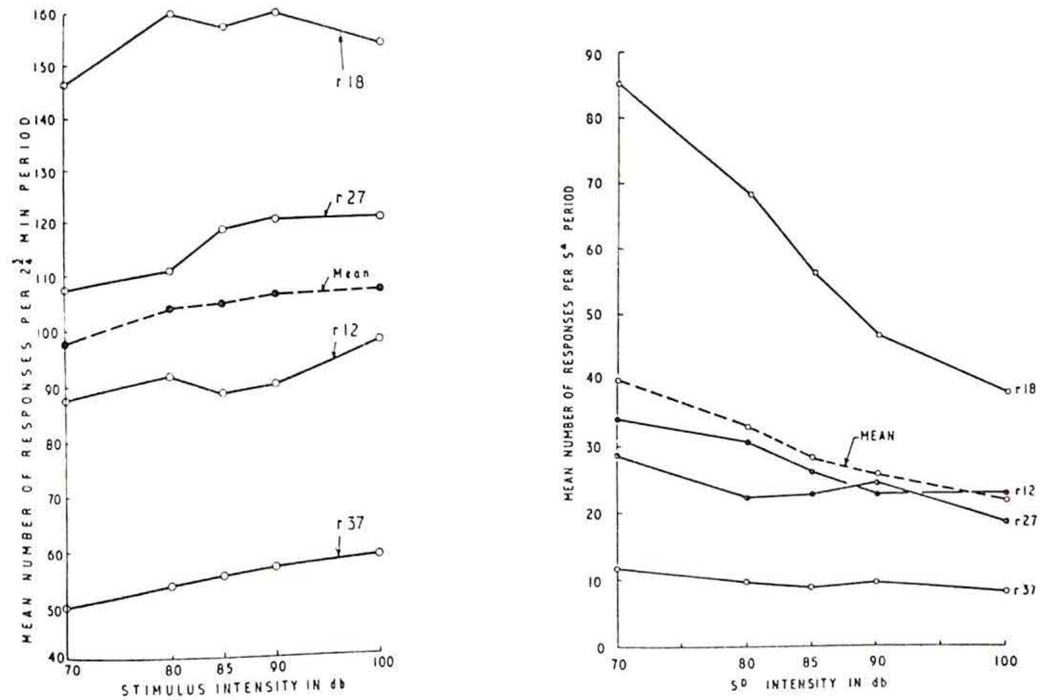


FIG. 5. Response rate as a function of S^D intensity in the Discrimination group in Exp. II: left-hand graph, S^D response rates; right-hand graph, S^Δ response rates.

TABLE 5

Effects of S^D Intensity on S^D and Subsequent S^Δ Response Rates in Exp. II: Significance Levels and Linear Regression Equations

S	Main Effect *F*	Linear Regression *F*	Departures from Linearity *F*	Regression Equation
		S^D Response Rates		
r 27	4.75**	15.62***	1.13	$y = 70.1 + 0.545x$
r 37	4.65**	18.11***	0.17	$y = 30.5 + 0.284x$
r 18	2.67*	2.65	2.68	—
r 12	7.36***	19.39***	3.35*	$y = 63.6 + 0.327x$
Group	8.85***	31.61***	1.26	$y = 74.5 + 0.353x$
		S^Δ Response Rates		
r 27	2.34	6.39*	0.98	$y = 46.7 - 0.269x$
r 37	1.41	4.25*	0.46	$y = 18.0 - 0.098x$
r 18	11.13***	40.61***	0.81	$y = 192.9 - 1.561x$
r 12	4.40**	15.16***	2.80	$y = 61.8 - 0.402x$
Group	23.62***	93.25***	0.41	$y = 80.6 - 0.594x$

Note.—$df = 4$ for the main effect, 1 for linear regression, and 3 for departures from linearity; the error variances have $df = 48$ in the case of the individual *Ss* and $df = 192$ in the case of the group analyses. In the regression equations, y = mean number of responses in a single period lasting 2 min., 45 sec. and x = noise intensity in db.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

S^D response rate and the negative effect of S^D intensity on subsequent S^Δ response rate observed in Exp. I. In general, the data suggest that both relationships are linear over the range of S^D intensities studied. However, linear regression of S^D response rate on S^D intensity is no longer significantly different from chance in the case of r 18. In contrast, linear regression of S^Δ response rate on preceding S^D intensity is now significant in the case of r 37, although it was not significant in Exp. I.

An effect of drive on the law of strength (predicted by Pavlovian theory) would be expected to appear in the present experiments in the form of an interaction between Drive and Intensity in the analyses of S^D responses. However, this interaction was not significant in any of these analyses, whether of the individual *Ss* or of the group as a whole. Indeed, only one interaction of any kind involving Intensity was significant. This was an interaction between Intensity and Replication (i.e. first 2 wk. of testing vs. second 2 wk.) in the case of r 18, $F(4, 48) = 4.53$, $p < .01$. Further analysis of this interaction by t tests showed that, in the first fortnight, rate at 100 db. was

TABLE 6

Coefficients of Linear Regression of S^D Response Rate on S^D Intensity and of S^Δ Response Rate on Preceding S^D Intensity as a Function of Drive in Exp. II

S	S^D Response Rate		S^Δ Response Rate	
	Low Drive	High Drive	Low Drive	High Drive
r 27	+0.609	+0.481	−0.237	−0.301
r 37	+0.339	+0.229	−0.022	−0.174
r 18	+0.292	+0.226	−1.603	−1.613
r 12	+0.403	+0.250	−0.306	−0.498

significantly lower than at either 85 or 90 db.; in the second fortnight, rate at 70 db. was significantly lower than at any other intensity; at all intensities, rate during the second fortnight was higher than rate during the first fortnight.

In spite of the nonsignificant Drive × Intensity interactions, the slopes of the curves relating response rate to S^D intensity were calculated separately for each drive condition in an effort to see whether, after all, any systematic changes had been induced by the change in drive level. The results of these calculations are presented in Table 6. It will be seen that, in the case of all four *S*s, the slope of the curve relating S^D response rate to S^D intensity is lower in the high-drive condition and the slope relating S^Δ response rate to S^D intensity is higher in the high-drive condition. Figure 6 presents the relevant curves for S^D responding, together with the corresponding curves from Exp. I.

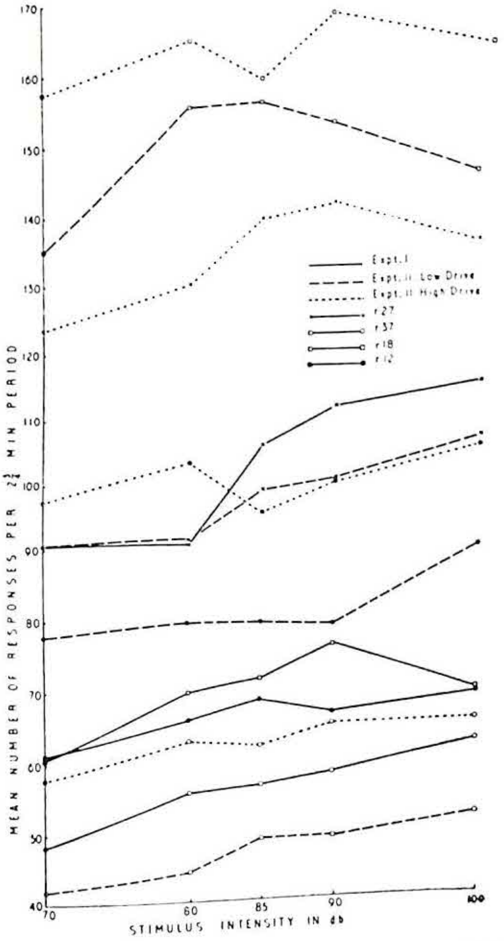


FIG. 6. S^D response rates as a function of S^D intensity and drive.

Effects of Noise Intensity and Drive in Subject r 20

Although noise intensity did not affect r 20's response rate in Exp. I, a significant Intensity effect appeared in Exp. II, $F\ (4, 48) = 3.79,\ p < .01$. This was accompanied by a significant interaction between Intensity and Replication, $F\ (4, 48) = 3.00,\ p < .05$. Further analysis of this interaction by t tests showed that the effect of noise intensity was almost entirely due to a very high response rate during 100-db. periods in the first fortnight only; there were no significant differences between intensities in the second fortnight. There was no interaction between Drive and Intensity.

DISCUSSION

The most important result to emerge from both Exp. I and Exp. II is that the intensity of a discriminative stimulus has a definite and very consistent effect on operant response rate. If we consider the four *S*s as four replications of each experiment, then in all eight replications a significant effect of S^D intensity on operant response rate was found, and, in all but one case, this effect was of the same kind: as stimulus intensity increased, so did response rate. It is not possible to be quite so definite about the exact shape of the function relating S^D intensity and operant response rate, but most frequently it has been monotonic and linear when stimulus intensity is

expressed in decibels. This was the finding in five of the eight replications.

Although the effect of S^D intensity on response rate in the present experiments was highly consistent, it should be noted that, quantitatively, it is not a very marked phenomenon: the data gained in these experiments suggest that, on the average, a rise of 10 db. in the intensity of the S^D is sufficient to increase response rate by only about 1.5 responses a minute. It is doubtful whether the customary scrutiny of cumulative records, without the aid of the kind of statistical tool which has been used in the present study, would have been able to establish the reality of this effect.

A very different picture emerges when we consider the effects of nonsignal noise intensity. The most frequent finding has been that response rate is unaffected by the intensity of nonsignal noise. Such effects as did appear were sporadic, occurring in some *S*s but not in others, in the group as a whole but not in any single *S*, or in some weeks but not in others. They were also irregular, showing little systematic trend on the occasions when they did occur and bearing very little resemblance either to the effect of S^D intensity or, indeed, to one another. Thus, r 9 in Exp. I (Fig. 2) presented a negative relationship between noise intensity and response rate, r 10 in the same experiment (Fig. 2) and the group of three *S*s exposed to alternating noise and silence (Fig. 3) showed a curvilinear relationship, and r 20 in Exp. II showed a positive relationship entirely due to a very high response rate at 100 db. in one fortnight only.

It seems safe to conclude, then, that the law of strength, or stimulus intensity dynamism, operates only in the case of discriminative stimuli. It should, however, be pointed out that this conclusion is not entirely beyond criticism. As well as the difference in the signal value of the stimulus, the treatments accorded the Discrimination and Nondiscrimination groups also differed in other ways which might conceivably be of importance. In particular, there was no interruption of VI reinforcement by extinction components for the Nondiscrimination group, and therefore rate of reinforcement was twice as great for this group.

The data allow us to discount other differences between the groups. Thus the slight difference in drive level in Exp. I (since the Nondiscrimination group was always tested after the Discrimination group) cannot account for the absence of a law of strength with nonsignal noise, in view of the demonstration in Exp. II that much greater differences in drive level may alter the slope of the stimulus intensity curve but do not alter its shape. Similarly, the slightly higher response rate recorded by the Nondiscrimination group in Exp. I (Fig. 2) cannot be of importance since the Discrimination group still displayed a law of strength with much higher response rates in Exp. II.

The fact that stimulus intensity dynamism appears only with signal stimuli is in agreement with Perkins' (1953) hypothesis accounting for this phenomenon in terms of generalization of inhibition from the unreinforced zero-intensity stimulus. However, the importance of the signal function of the stimulus could be due to other causes than those postulated by Perkins. Pavlovian theory also suggests that only signal stimuli are concerned in the law of strength. A further possibility stems from recent work on the neurophysiological processes involved in habituation (Hernández-Peón, 1961; Sharpless & Jasper, 1956): it could be suggested that the intensity of a nonsignal stimulus fails to affect the magnitude or rate of conditioned responding because it is either blocked before it reaches the CNS, or because it does not get sufficient "boost" from the reticular arousal system to affect cortical functioning. More convincing support for Perkins' hypothesis comes from the effect of S^D intensity on subsequent S^Δ response rate observed in the present experiments. The most likely explanation for this effect (also probably confined to signal stimuli) is that it is due to generalization of *excitation* from low intensities of the positive stimulus to the zero-intensity

negative stimulus. This explanation, together with Perkins' account of stimulus intensity dynamism, leads us to expect that the positive effect of S^D intensity on S^D response rate and the negative effect of S^D intensity on S^Δ response rate will occur together (as they did in these experiments), since both arise from the same failure to discriminate between positive and negative stimuli.

The failure of any of the interactions between Drive and Intensity (Exp. II) to achieve statistical significance is disappointing from the standpoint of Pavlovian theory (Gray, 1964). This predicts that an increase in drive should produce either a curvilinearity in the curve relating response strength to stimulus intensity or an asymptote at higher stimulus intensities. However, the experiments did produce evidence that something in the nature of the latter process was occurring, since all four *S*s displayed a lower slope of the stimulus intensity/response rate curve in the high-drive condition. Other evidence in favor of Pavlovian theory comes from a consideration of the results achieved with r 18. This *S* displayed a much greater rise in response rate from Exp. I to Exp. II than from the low-drive to the high-drive condition of Exp. II (Table 1). Assuming that this represents an increase in drive (due, no doubt to the reduction in h_t), it is in agreement with Pavlovian theory that linear regression of response rate on S^D intensity was no longer significant under the higher drive level of Exp. II in this *S*.

A consideration of the individual differences observed in these experiments also offers some support for Pavlovian theory. According to this theory (Gray, 1964), the absence of a law of strength, or the presence of curvilinearities in the law of strength curve such that highest response strength is reached at some point below maximum stimulus intensity, are signs of "weakness of the nervous system." Such signs were shown in both experiments by r 18, suggesting that they are a stable feature of this individual's responding. Some indication of similar tendencies appears in the high-drive curve obtained from r 27 in Exp. II (Fig. 6). Thus, of the four *S*s studied, r 18 could be judged the "weakest" and r 27 the next weakest. Now, there are grounds for believing that "weak" animals are individuals which are particularly easily stimulated to the point of giving their maximum possible response (Gray, 1964). It is therefore consistent with the predictions of Pavlovian personality theory that these two *S*s had much higher response rates than r 37 and r 12 (means for Exp. I and II combined: r 18—41.0 responses/min; r 27—39.9; r 12—28.6; r 37—20.1). Considering the law of strength as a joint function of drive and individual differences (as Pavlovian theory demands), it can be seen from Fig. 6 that the irregularities in this effect become much more marked as response rate increases and that the nature of these irregularities is, generally speaking, of the expected kind.

REFERENCES

Barnes, G. W. Conditioned stimulus intensity and temporal factors in spaced-trial classical conditioning. *J. exp. Psychol.*, 1956, **51**, 192–198.

Beck, S. B. Eyelid conditioning as a function of CS intensity, UCS intensity, and manifest anxiety scale score. *J. exp. Psychol.*, 1963, **66**, 429–438.

Berlyne, D. E. Stimulus intensity and attention in relation to learning theory. *Quart. J. exp. Psychol.*, 1950, **2**, 71–75.

Broadhurst, P. L. Applications of biometrical genetics to the inheritance of behaviour. In H. J. Eysenck (Ed.), *Experiments in personality.* Vol. 1. *Psychogenetics and psychopharmacology.* London: Routledge & Kegan Paul, 1960. Pp. 1–102.

Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.

Carter, L. F. Intensity of conditioned stimulus and rate of conditioning. *J. exp. Psychol.*, 1941, **28**, 481–490.

Eisman, E. An investigation of the parameters defining drive (*D*). *J. exp. Psychol.*, 1956, **52**, 85–89.

Grant, D. A., & Schneider, D. E. Intensity of the conditioned stimulus and strength of

conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.

GRAY, J. A. (Ed.) *Pavlov's typology: Recent theoretical and experimental developments from the laboratory of B. M. Teplov.* Oxford: Pergamon Press, 1964.

HERNÁNDEZ-PEÓN, R. Reticular mechanisms of sensory control. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 497–520.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.

HOVLAND, C. I. The generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–290.

JAY, G. E., JR. Genetic strains and stocks. In W. J. Burdette (Ed.), *Methodology in mammalian genetics.* San Francisco: Holden-Day, 1963. Pp. 83–126.

KAMIN, L. J., & SCHAUB, R. E. Effects of conditioned stimulus intensity on the conditioned emotional response. *J. comp. physiol. Psychol.*, 1963, **56**, 502–507.

KESSEN, W. Response strength and conditioned stimulus intensity. *J. exp. Psychol.*, 1953, **45**, 82–86.

KIMMEL, H. D. Amount of conditioning and intensity of conditioned stimulus. *J. exp. Psychol.*, 1959, **58**, 283–288.

KIMMEL, H. D., HILL, F. A., & MORROW, M. C. Strength of GSR and avoidance conditioning as a function of CS intensity. *Psychol. Rep.*, 1962, **11**, 103–109.

MILLER, W. C., & GREENE, J. E. Generalization of an avoidance response to varying intensities of sound. *J. comp. physiol. Psychol.*, 1954, **47**, 136–139.

MYERS, A. K. Effects of CS intensity and quality in avoidance conditioning. *J. comp. physiol. Psychol.*, 1962, **55**, 57–61.

PASSEY, G. E., & HERMAN, P. N. The shape of the discrimination gradient for two intracontinuum stimulus separations. *J. exp. Psychol.*, 1955, **49**, 273–277.

PASSEY, G. E., & POSSENTI, R. G. The effect of conditioned stimulus intensity upon a simple running response. *J. genet. Psychol.*, 1956, **89**, 27–33.

PERKINS, C. C., JR. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225–231.

RAZRAN, G. The dominance-contiguity theory of the acquisition of classical conditioning. *Psychol. Bull.*, 1957, **54**, 1–46.

SHARPLESS, S., & JASPER, H. Habituation of the arousal reaction. *Brain*, 1956, **79**, 655–680.

WALKER, E. G. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1960, **59**, 303–311.

WILCOTT, R. C. A search for sub-threshold conditioning at four different auditory frequencies. *J. exp. Psychol.*, 1953, **46**, 271–276.

ZENER, K. Individual differences in the functional relationship of conditional stimulus intensity and response magnitude. *Amer. Psychologist*, 1949, **4**, 218. (Abstract)

(Received November 1, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 25–29

ACQUISITION AND EXTINCTION AFTER INITIAL TRIALS WITHOUT REWARD [1]

NORMAN E. SPEAR

Rutgers University

WINFRED F. HILL AND DENIS J. O'SULLIVAN

Northwestern University

The effect of nonrewarded trials which preceded any rewarded trials was studied with rats in a straight runway. Running speeds were measured for groups given either 24 nonrewarded trials preceding 24 rewarded (24N-24R), 24 nonrewarded trials interspersed among 24 rewarded (50%), 24 rewarded trials (24R), or 48 rewarded trials (48R). At the end of acquisition, the mean speeds were ordered as: 50% > 48R > 24R > 24N-24R, with Group 24N-24R showing marked forgetting on the first trial of each day after reward was introduced. During extinction the ordering was: 50% > 48R > 24R > 24N-24R on the first trial of each extinction day, but 50% > 24N-24R > 48R = 24R on the last trial of each extinction day.

The present experiment investigated the effect of nonrewarded trials *prior* to the introduction of reward on running speed during subsequent acquisition and extinction. In addition, the effects of these initial nonrewarded trials were compared with the effects of an equivalent number of nonrewarded trials which were interspersed among rewarded trials.

METHOD

General design.—The experiment included four groups, each of which was given an acquisition and an extinction session in a straight-alley runway. Of basic interest was the comparison made between the extinction performance of a group given 24 nonrewarded trials in the runway followed by 24 rewarded trials (Group 24N-24R), and a group given 48 trials, all of which were rewarded (Group 48R). In order to establish the effects of the nonrewarded trials per se, a third group was also included which received only 24 rewarded trials (24R). The fourth group was a standard partial-reinforcement group which received 24 nonrewarded trials interspersed among 24 rewarded trials (Group 50%). Following acquisition training, 8 trials of extinction were given per day for 4 successive days.

Subjects and apparatus.—The *S*s were 60 (15 per group) naive female albino rats of the Sprague-Dawley strain, 90 days old at the start of runway training. The apparatus was the 4-ft. enclosed runway described by Lewis (1956). In this runway, running time is measured from the opening of the start-box door until *S* enters the goal box.

Procedure.—Prehandling and feeding schedule were similar to that employed by Hill and Spear (1962), except *S*s were (*a*) prehandled *without* access to food pellets and (*b*) never presented with food sooner than 15 min. after *S* had been handled by *E*. These modifications were introduced to separate the eating and handling experience prior to runway training.

Eight *S*s, two from each of the four groups, were run in rotation with a minimum interval of 3 min. between successive trials (eight per day for each *S*). During the first 3 days of training, *S*s in Group 24R were handled near the runway for a period of time equivalent to that experienced by the other *S*s. They were not, however, run in the runway until the rewarded trials of Group 24N-24R began on Day 4.

The schedule of reinforcement was the same for all *S*s in Group 50% and was modified from that employed by Wagner (1961). The daily sequences were as follows: +−+−−++−, −++−−−++, +−−++−+−, −++−−++−, −−+++−−+, +−+−−−++.

[1] This research was supported by Grant G-8706 from the National Science Foundation.

During all trials, the start-box door was opened as soon as *S* was oriented toward it. Following a run, *S* was confined in the goal box for 10 sec. On rewarded trials, two .045-gm. Noyes pellets were present in a metal food cup in the goal box. The food cup was removed on nonrewarded trials.

Results and Discussion

The course of acquisition is shown in Fig. 1. Group 24R which received only 24 trials, all rewarded, is shown on the graph from Trials 25 to 48.

Three facts are of particular interest with respect to these acquisition speeds. (*a*) Group 24N-24R, which received 24 nonrewarded trials preceding its 24 rewarded trials, demonstrated marked forgetting on the initial trials of the second and third days of acquisition, and at no time gave any indication of latent learning. (*b*) When the average speeds of these two groups were compared by Duncan's multiple-range test (Edwards, 1960) on the last day of acquisition, Group 24R was found to have run faster than Group 24N-24R ($p < .05$). It is clear, however, that this difference was due to the forgetting exhibited by Group 24N-24R and reflected in the first few trials of the day. (*c*) Group 48R, which was given 48 trials, all rewarded, was considerably faster ($p < .001$) at the end of acquisition than either of Groups 24N-24R or 24R. It was, however, slower than Group 50% ($p < .001$). This latter finding is consistent with the results of several investigators (see Spence, 1960).

Extinction curves for the four groups are shown in Fig. 2. The vertical lines represent a 24-hr. interval between trials. It is of interest to note how the speed of Group 24N-24R, whose first 24 runway trials were not rewarded, dips after a 24-hr. interval as it did in acquisition. Per-

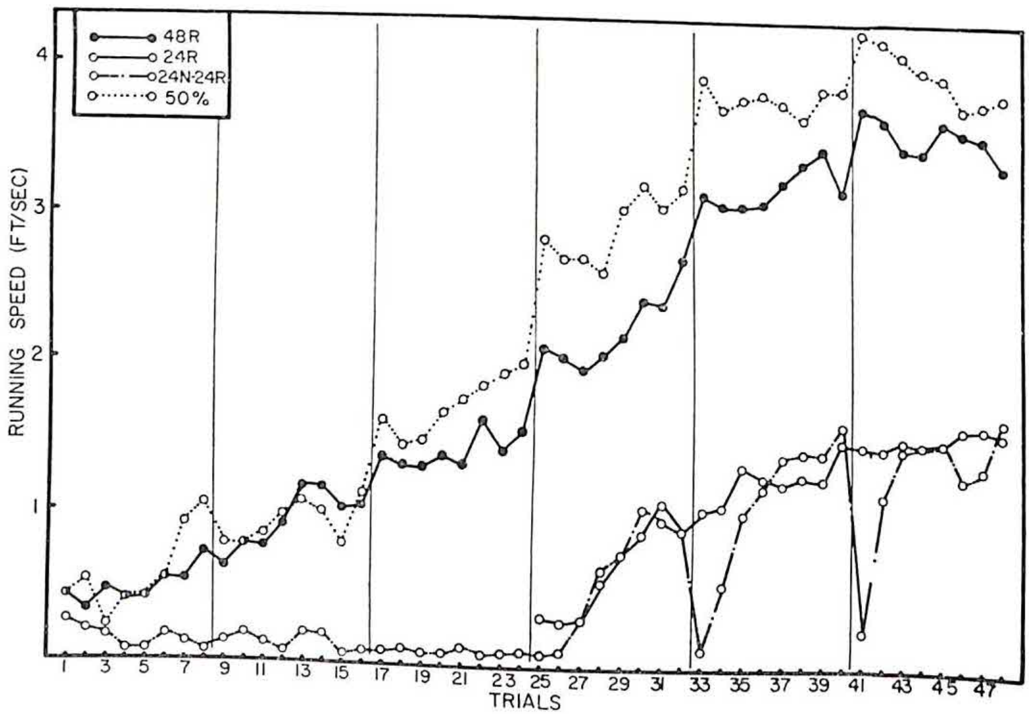


FIG. 1. Mean running speed for each acquisition trial. (The vertical lines represent a 24-hr. interval between trials.)

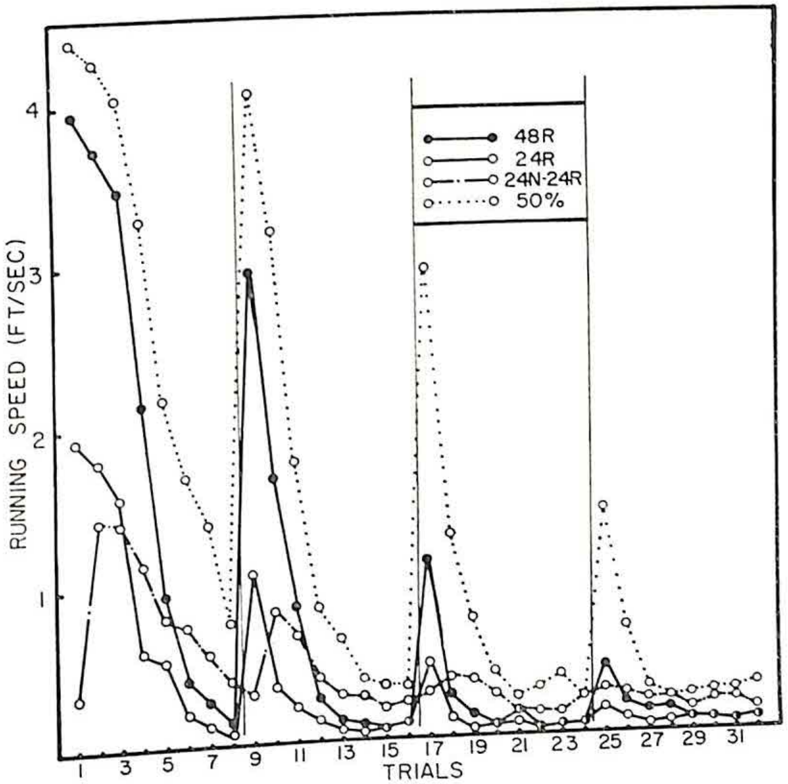


FIG. 2. Mean running speed for each extinction trial. (The vertical lines represent a 24-hr. interval between trials.)

haps more important is the *increase* in speed from the first to the second trial of a day even though the first trial was not rewarded. This pattern is clearly different from that of the other three groups which all showed marked spontaneous recovery on the first trial of each day and then decreased consistently. Because of these sharp and interesting differences in the pattern of running within a day, comparisons were made of speed on both the first and the last trials of each extinction day, and also in terms of trials to an extinction criterion.

A Groups × Days analysis of variance was performed on daily first-trial speeds in extinction. As would be expected, the overall effect of groups was significant, $F\ (3, 56) = 289.95$, $p < .001$, as was the decrease in speed from day to day, $F\ (3, 168) = 331.29$, $p < .001$. In addition, the interaction of days with groups was significant, $F\ (9, 168) = 52.33$, $p < .001$. This interaction undoubtedly reflected the effect of the truncated distribution of speeds in extinction, but the lack of a speed change from day to day within Group 24N-24R contrasted with the consistent drop of the other groups, and was also a component of this interaction. It is clear in Fig. 2 that first-trial speed of this group was essentially constant throughout extinction. Individual comparisons of the overall group means by Duncan's multiple-range test showed that all differences were significant at the .001 level, with an ordering from fast to slow speeds of: 50%, 48R, 24R, 24N-24R.

As can be seen in Fig. 2, the relationships of the groups changed rather

rapidly within a day after the first trial. In particular, Groups 24R and 48R that had never previously received nonrewarded trials dropped below the other groups by the end of the day. In terms of speed on the last trial of each day, a Groups × Days analysis of variance revealed that the effects of groups, F (3, 56) = 24.18, $p < .001$; days, F (3, 168) = 11.45, $p < .001$; and their interaction, F (9, 168) = 3.34, $p < .001$, were all significant. Comparison of speeds on the last trial of each extinction day summed over days (Duncan's test) showed that the mean speed of Group 50% exceeded that of Group 24N-24R ($p < .001$), and that both of these were faster than Groups 24R and 48R ($p < .001$). There was no difference between the latter two groups. These same relationships prevailed with a post-hoc trials-to-criterion measure. A criterion of two successive "slow" trials with traversing time greater than 20 sec. was employed since many *S*s in Groups 50% and 24N-24R never attained a more stringent criteria. Employing Duncan's multiple-range test, it was found that Group 50% took more trials to extinction than Group 24N-24R ($p < .05$), which in turn took more ($p < .01$) than the other two groups, which did not differ.

Finally, the number of trials on which an *S* did not enter the goal box within the maximum allotted time of 2 min. should be mentioned. Here again the same relationships prevailed. Every *S* within Groups 24R and 48R made at least one such nonrun, but only 4 of the 15 *S*s in Group 24N-24R and no *S*s in Group 50% made any nonruns.

Thus, the two major sets of facts obtained with this experiment concern the effect of initial nonrewarded trials on (Fact 1) daily first-trial running speed after the introduction of reward, and (Fact 2) running speed in subsequent extinction. These will be discussed in order.

1. Hungry *S*s, whose initial runway experiences did not include food reward, ran at their initial speed (relatively slow) early on each day even after reward was introduced. This regression to earlier speeds was quite transient within days and disappeared after the first few trials. Between days, however, this regression persisted through 3 days of rewarded trials, and was still evident during the first 2 extinction days. It is suggested that this phenomenon be viewed as greater forgetting between days by Group 24N-24R, relative to the other groups, due to the recovery of habits established during the initial nonrewarded trials. The recovery of the slower response that had been appropriate on earlier trials interfered with the faster speed of running which was more consistent with the later current reward. Thus, the reduced speed by Group 24N-24R on the first trial of each day after reward was introduced is viewed within the context of the traditional proactive interference paradigm. This interpretation is consistent with the results of Collier, Knarr, and Marx (1961) who studied the effects on running speed of shifts in sucrose-concentration reward. These writers (Collier et al., 1961) note the suggestion from their data that, ". . . The rat starts each day at a level commensurate with the previous reinforcement [p. 494]." Hill and Spear (1963) have also reported this phenomenon.

2. Nonrewarded trials, which were given prior to a series of trials in which food reward was presented for the first time, had the effect of increasing asymptotic extinction speed and of extending the number of trials to a criterion of extinction. The importance of the patterning of the nonrewarded trials was also evident, since even greater asymptotic extinction speed and more trials to an extinction criterion were found when rewarded trials were interspersed among

the nonrewarded trials. It should be clear that the conclusion regarding asymptotic speed depends upon exclusive consideration of trials occurring late in an extinction day. Furthermore, it is not inconceivable that the differences at asymptotic extinction might have disappeared with more extinction trials. Nevertheless, the trials-to-criterion data and the shifts of group ordering from acquisition to extinction seem to compensate for such measurement problems.

The apparently increased resistance to extinction of Group 24N-24R would appear to create interpretation problems for certain theories of the partial-reinforcement extinction effect. For example, Amsel (1958) clearly requires that *S* must have built up a substantial expectancy (i.e., sufficiently strong r_G-s_G) *before* experiencing the nonreward in acquisition, if resistance to extinction is to be increased. Lawrence and Festinger's view, though not completely unambiguous, apparently does not make this requirement (Lawrence & Festinger, 1962, pp. 113–115). However, Lawrence and Festinger's concepts would need to be extended to account for the superior extinction performance of Group 50%.

There appears to be no obvious satisfactory explanation for the effect of the initial nonrewarded trials on extinction behavior. It is likely that such an explanation will be found only after more is known about the precise nature of *S*'s behavior during the initial nonrewarded trials. Perhaps it is not proper to consider the running behavior of Group 24N-24R within the same context of "resistance to extinction" as is employed elsewhere.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situation. *Psychol. Bull.*, 1958, **55**, 102–119.

Collier, G., Knarr, F. A., & Marx, M. H. Some relations between the intensive properties of the consummatory response and reinforcement. *J. exp. Psychol.*, 1961, **62**, 484–495.

Edwards, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

Hill, W. F., & Spear, N. E. Resistance to extinction as a joint function of reward magnitude and the spacing of extinction trials. *J. exp. Psychol.*, 1962, **64**, 636–639.

Hill, W. F., & Spear, N. E. Reminiscence and forgetting in a runway. Paper read at Psychonomic Society meetings, 1963.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcement: The psychology of insufficient reward.* Stanford, Calif.: Stanford Univer. Press, 1962.

Lewis, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial interval. *J. exp. Psychol.*, 1956, **51**, 45–53.

Spence, K. W. *Behavior theory and learning: Selected papers.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Wagner, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

(Received November 22, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 30–34

INFLUENCE OF RESPONSE DISCRIMINABILITY ON STIMULUS DISCRIMINABILITY[1]

J. F. RICHARD

University of Rennes, France

2 experiments tested the following hypothesis: if 2 stimuli are equivalent if they are followed by the same event, or by 2 physically different but not discriminable events, we can expect that the differentiation of a set of stimuli will be poorer if they are associated with events (or responses) of low discriminability. Experiment I used response cards showing varying number of dots, in high to low discriminability condition, associated with circles of varying diameter. In Experiment II, rectangles of varying height were associated with motor responses. Results of these experiments led to the conclusion that the discriminability of a set of stimuli depends on the discriminability of the responses with which they have been previously associated.

Several studies have shown that two stimuli become less discriminable if they are associated with the same response; this effect is called acquired equivalence of cues (De Riviera, 1959; Dollard & Miller, 1950; Grice & Davis, 1958, 1960; Jeffrey, 1953; Kurtz, 1955; Lacey, 1961). This principle may be applied when responses are not easily discriminable. We may assume that in this case *S* will not perceive as many responses as there are in fact. For instance, if there are eight different responses, *S* may perceive only five or six different ones. Let us suppose that eight stimuli are to be associated with these eight responses; from *S*'s point of view it is as if these stimuli were to be associated with five or six responses. We then have a situation in which a set of stimuli is associated with a smaller set of responses, the typical situation for acquired equivalence of cues. Therefore, we can expect that decreasing response discriminability will have the same effect as reducing the number of responses—that is, a decrease of stimulus discriminability. Such a result would mean that stimulus discriminability is dependent on response discriminability. Two experiments have been conducted to test this hypothesis.

[1] This paper is a part of a "Doctorat de Troisième Cycle" thesis (Richard, 1963) presented by the author at the University of Paris, May 1963. This research has been conducted under the supervision of P. Fraisse, to whom we are grateful for his helpful comments. The experiments have been conducted at the Laboratoire de Psychologie of the University of Rennes.

Experiment I

Method

Materials.—The stimuli used were eight black circles on white cards. Diameters of the circles varied from 40 to 64 mm. in a logarithmic progression. On the response cards sets of dots were drawn at random in a matrix of 64 cells. Cards with 8, 11, 14, 18, 24, 32, 42, and 55 dots were used in the high discriminability response condition, and cards with 20, 22, 24, 26, 29, 32, 35, and 38 dots were used in the low discriminability response condition.

Stimulus and response cards were presented by hand through an opening in the center of a board. Presentation time was about 10 sec.

In both conditions circles of increasing size were associated with response cards with decreasing numbers of dots in perfect correlation.

Procedure.—The *S*s were told that they were going to see circles of different size

followed by cards with varying numbers of dots on them. The circles were supposed to represent marksmen and the cards the target at which they had fired. The *S*s had to judge the marksmen's skill from the number of dots on the target, each dot representing a successful shot. The *S*s also had to rate the marksmen by skill.

The experiment began by presenting each circle twice at the same time as the respective target in a random order. Afterwards the circle was presented first and *S* wrote his response, then the target was presented. The experiment was continued for eight trials, each stimulus and each response being presented eight times. At the end of the experiment *S*s were asked how many circles they thought there had been.

The experiment was carried out in a group. The *S*s were college students, 16 in each condition.

Results

We shall analyze three types of data: (*a*) judgments about number of circles used in the experiment; (*b*) number of different categories used to indicate the position of the marksmen; (*c*) transmitted information as a measure of stimulus discriminability.

The median of the distribution of the estimated number of circles is 6.8 in the high discriminability response condition and 5.8 in the low discriminability response condition. A median test shows this difference to be significant beyond the .05 level. The *S*s in the low discriminability response condition have thus perceived on average a circle less than those in the high discriminability response condition.

For the number of categories used to rate the marksmen, the corresponding values are 7.0 and 6.5. This difference is consistent with the previous one but is not statistically significant.

We have now to see if stimulus discriminability is lower when responses are weakly discriminable. We have computed transmitted information for each set of *S*s which used the same number of categories in rating the marksmen in each condition. Since the number of dots on the cards is perfectly correlated with the size of the circles, the task of indicating the marksmen's position as regards their skill is in fact a task of identification of the stimuli. Some *S*s were dropped because too few had used the same number of categories. The data of the last four trials have been used for this computation. As is shown by Table 1, transmitted information is higher in the high discriminability response condition than in the low discriminability response condition. As the value obtained for each category provides an independent estimate of transmitted information, the probability of such an arrangement of the transmitted values in the two conditions may be calculated as follows: if we call A each estimate of T in the high response discriminability condi-

TABLE 1

TRANSMITTED INFORMATION FOR EACH CONDITION

	No. of Categories Used	No. of *S*s	Transmitted Information	Weighted *M*
High discriminability response cond.	6 7 8	3 9 2	1.62 1.93 1.98	1.87
Low discriminability response cond.	5 6 7	4 5 3	1.25 1.57 1.38	1.42

tion and B each estimate of T in the low response discriminability condition, the probability of getting in a random sampling the arrangement AAABBB of the values in order of magnitude is $3!3!/6! = .05$. Stimulus discriminability is then lower when stimuli have been associated with poorly discriminable responses than when they have been associated with highly discriminable ones.

Experiment II

In the preceding experiment *S*s were informed visually on stimulus-response assignment. In many cases, however, this information is dependent on response production and is given by kinesthetic sensitivity. In Exp. II response discriminability is varied by the pattern of motor response required and, therefore, by the pattern of kinesthetic feedback.

Method

Apparatus.—The stimuli were eight rectangles 20 mm. wide and varying in height from 40 to 68 mm. in equal steps of 4 mm. They were presented by slides by means of an automatic projector. Presentation time was 7.5 sec.

The response display was adapted from Shepard's (1958) experiment. Eight 3-mm. metal contacts were arranged in a row along the bottom of a box. The *S* was required to insert an electric stylus into a narrow rectangular slot in the top of the box. Contacts were inserted in the bottom of the box in such a way that the spaces between the contacts were at the same level as the contacts themselves, and *S* was unable to distinguish individual contacts by tactual cues. Contacts were 12 mm. apart in the high discriminability response condition and 4 mm. apart in the low discriminability response condition. A panel was arranged in front of the apparatus, so that *S* could not see the display, particularly the slot and the stylus. There were, therefore, no visual cues available for the control of the response.

On each presentation *S* had to try to find the correct contact; a red lamp in the top of the box informed *S* when he had succeeded. Each rectangle was associated with a contact; contacts from left to right corresponded with rectangles of, respectively, decreasing size. The *E* closed the circuit for the required contact when the latter was touched by the stylus; when each rectangle was presented *E* placed the appropriate switch at the contact position.

Procedure.—The *S*s were told they were going to see different rectangles and had to find the position which corresponded to each of them at the bottom of the box; they were asked to remember the position so that they might find it again when the rectangle was presented a second time.

During the learning period each stimulus was presented 12 times in a random order. This was followed by an identification task using the absolute judgment method, but *S* had to choose himself the number of categories to be used. Each stimulus was presented three times.

Beside the high discriminability response condition and the low discriminability response condition a control condition was added, in which *S*s were given only the identification task without preliminary training.

The *S*s were 37 manual workers, 13 in each motor condition and 11 in the control condition.

Results

We shall consider at first the number of categories used in the identification task, which is an estimation of the number of different rectangles each *S* has perceived. The median of the distribution is 6.0 in the control condition, 5.6 in the high discriminability response condition, and 4.3 in the low discriminability response condition. The difference between the two latter conditions, as tested by the Mann-Whitney *U* test, is significant beyond the .05 level.

Transmitted information has been computed for the groups of *S*s having used the same scale of judgment in each condition: these values are presented in Table 2 as well as the mean weighted index for each condition. The values for the high discriminability response condition are consistently higher than those for the

TABLE 2

TRANSMITTED INFORMATION FOR EACH CONDITION

	No. of Categories Used	No. of *Ss*	Transmitted Information	Weighted *M*
Control cond.	5	2	1.45	1.46
	6	3	1.38	
	7	3	1.54	
High discriminability response cond.	4	4	1.38	1.38
	5	2	1.42	
	6	4	1.36	
Low discriminability response cond.	3	2	0.93	1.09
	4	3	1.13	
	5	4	1.14	

low discriminability response condition. The probability of getting such an arrangement of the values in a random sorting, calculated as above is .05. We may then say that the discriminability of the rectangles has been lowered by their association with weakly discriminable motor responses.

Discussion

The results of these experiments are consistent. Stimuli are less well discriminated when they have been associated with weakly discriminable responses. This is a confirmation of our hypothesis: decreasing response discriminability has the same effect as reducing the number of responses associated with stimuli and consequently discriminability is dependent on response discriminability.

While many experiments have investigated the effect of mediated equivalence of cues by having the same response associated to several stimuli, very few studies have used similar instead of identical responses. We know only two studies in which this paradigm was used, one by Dietze (1955) and one by Norcross (1958). In Dietze's experiment children had to learn either similar (BEEM, MEEM, PEEM) or dissimilar (JOD, DAF, MEEP) words in response to shapes. The results show that shapes are easier to discriminate when they have been associated with dissimilar words than when they have been associated with similar ones. In Norcross' study children had to pair two similar (ZIM, ZAM) and two dissimilar (WUG, KOS) nonsense syllables with four photographs and were then required to press one of four keys in response to each stimulus. More mistakes were made for the pair of stimuli with which similar names had been associated than for the other. These results are consistent with ours: response similarity tends to increase stimulus similarity while response dissimilarity has the opposite effect. Le Ny (1959) has reported a similar effect: six tones of increasing pitch and equally spaced in jnd's were associated with six letters (A, B, C, D, E, and F, respectively, for the control group and A, B, C, U, X, Z for the experimental group). The *Ss* had then to rate the difference in pitch between each tone: the third and the fourth tone (in increasing order of pitch) were rated more distant by the experimental group (for which C and U had been associated with these tones) than by the control group (for which C and D had been associated with these tones). The distance between stimuli may then be modified by the distance between responses.

From these results it is possible to predict that if stimulus discriminability is dependent on response discriminability, then reducing stimulus similarity would have little effect when response similarity is high. It would be more beneficial to reduce response similarity

rather than stimulus similarity. In the same way, if stimulus similarity and response similarity are high, a greater facilitation effect would be expected from response predifferentiation than from stimulus predifferentiation.

REFERENCES

De Riviera, J. Some conditions governing the use of the cue-producing response as an explanatory device. *J. exp. Psychol.*, 1959, **57**, 299–304.

Dietze, D. The facilitating effect of words on discrimination and generalization. *J. exp. Psychol.*, 1955, **50**, 255–260.

Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

Grice, G. R., & Davis, J. D. Mediated stimulus equivalence and distinctiveness in human conditioning. *J. exp. Psychol.*, 1958, **55**, 565–571.

Grice, G. R., & Davis, J. D. Effect of concurrent responses on the evocation and generalization of the conditioned eyeblink. *J. exp. Psychol.*, 1960, **59**, 391–395.

Jeffrey, W. E. The effects of verbal and nonverbal responses in mediating an instrumental act. *J. exp. Psychol.*, 1953, **45**, 327–333.

Kurtz, K. H. Discrimination of complex stimuli: The relationship of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283–292.

Lacey, H. M. Mediating verbal responses and stimulus similarity as factors in conceptual naming by school age children. *J. exp. Psychol.*, 1961, **62**, 113–121.

Le Ny, J. F. Similitude conditionelle et illusion auditive après une association sons-lettres. *Annee psychol.*, 1959, **59**, 47–59.

Norcross, K. J. Effects on discrimination performance of similarity of previously acquired stimulus names. *J. exp. Psychol.*, 1958, **56**, 305–309.

Richard, J. F. Généralisation du signal et généralisation de la réponse. Unpublished doctoral dissertation, University of Paris, 1963.

Shepard, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55**, 509–523.

(Received November 26, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 35–39

ANAGRAM SOLUTION TIMES: A FUNCTION OF THE "RULEOUT" FACTOR [1]

ROYCE R. RONNING

University of California, Berkeley

Anagram solution may be considered to be a process in which an S permutes in an algorithmic fashion all letters in the anagram. However, Ss may find it possible to "rule out" certain permutations because attempting these permutations is highly unlikely to yield an English word. 20 adult Ss were presented with 20 5-letter anagrams in a 20 × 20 Latin square. Ss were randomly assigned to sequences. The anagrams were chosen from 2 levels of frequency on the basis of the Thorndike-Lorge (1944) count. At each frequency level 5 of the words were "high" ruleout, i.e., permitted elimination of 90–94 of the possible 120 permutations and 5 were "low" ruleout, permitting elimination of 70–76 of the permutations. Anagrams with "high" ruleout totals were solved significantly more rapidly ($p < .01$) at both frequency levels.

Solution word frequency has been established as a potent variable in predicting anagram solution times (Mayzner & Tresselt, 1958, 1959). However, even within a class of words of a designated frequency of occurrence considerable between-words variability in solution time remains. A number of hypotheses have been advanced to explain this variability; among these hypotheses are: letter order of the anagrams, transition probabilities (bigram frequencies) of the anagrams, and transition probabilities of the solution words.

Letter order has proved useful in predicting anagram solution (Hunter, 1959; Mayzner & Tresselt, 1958); however, no objective way of classifying the various letter orders possible appears to exist. In the studies cited, letter orders were classified as "easy" or "hard" on the basis of inspection of the similarity or dissimilarity of the anagram letter order to the solution-word letter order. Erlebacher (1962) used Kendall's tau to classify systematically the possible rearrangements of five-letter words. Her conclusions (p. 49) suggest that the similarity of the anagram letter order to the solution-word order is not relevant to speed of solution.

Transition probabilities based on bigram frequencies (Underwood & Schulz, 1960) are not independent of word frequency, since they reflect the occurrence of various pairs of letters (bigrams) which are most commonly used in the English language. As a consequence, introduction of transition probabilities as a variable may be expected to yield contradictory results (Beilin & Horn, 1962; Mayzner & Tresselt, 1959).

At its simplest level, solution of a five-letter anagram by *S* may be conceived of as a simple problem-solving task which consists of trial and error manipulation of the five letters to form a five-letter word. Since there are 120 (5!) possible permutations of the five letters, an algorithmic approach would inevitably result in success. However, solution attempts

[1] This research was supported in part by the Research Committee of the School of Education of the University of California, Berkeley. The writer is grateful to J. L. Byers and M. J. Slakter for helpful comments and critical reading of the manuscript.

of this type quickly result in the discovery that certain bigrams and trigrams hypothesized as the *beginning* of the solution word simply do not appear in English words. With the elimination of each of these letter sequences (and the possible permutations of the remaining letters) a large number of the 120 possible combinations may be ruled out as impossible or highly improbable solutions.

To illustrate, consider the anagram HIGTL. Among these letters the following bigrams and trigrams simply do not occur as the *beginning* sequence of letters in English words: HG, HT, HL, GT, TG, TL, LH, LG, LT, GHT, GHL, GLH, GLT, THL, and THG. Each bigram rules out six possible orders and each trigram rules out two orders. Thus logical analysis rules out 54 + 12 = 66 of the 120 possible permutations.

Rarely do five-letter words begin with a vowel. A count at two levels of word frequency (five-letter words in the first thousand and seventh thousand of the Thorndike-Lorge, 1944, count) revealed that less than 10% of the five-letter words in these 2,000 words begin with a vowel. If, then, beginning a word with a vowel is also ruled out, for the anagram HIGTL, 24 additional permutations can be eliminated, thus bringing the final total ruled out to 90 (66 + 24) of the 120 permutations possible.

Mayzner and Tresselt (1958, 1959, 1962) suggest that anagram solution is an S-R mediational type of problem solving. In their model the anagram evokes a "variety of implicit processes" from *S* which relate letter order, word frequency, etc., in such a way as to produce solution in varying lengths of time. The present study, while not inconsistent with the mediational approach, suggests the possibility of looking at these presumed implicit processes from the standpoint of *S*'s intuitive acquaintanceship with the structural properties of the English language. Thus, *S*, on exposure to an anagram, chooses for a provisional try two or three letters which, from his prior acquaintance with English words, can be expected to lead to an English word. Since Miller (1951) indicates that over 55% of words in the English language follow the form consonant-vowel (cv) or consonant-consonant-vowel (ccv), *S* is likely, following the present model, to choose as initial letters either cv or cc (to be followed by a vowel). This may explain Hunter's (1959) finding that vowel-beginning words form much more difficult anagrams than do consonant-beginning words.

Following the provisional try, *S* either succeeds (solves the anagram) or rejects the letter sequence and chooses another set of letters and repeats the process. From this line of reasoning it is clear, then, that anagrams having a relatively small number of two- or three-letter combinations (high ruleout) which potentially yield English words when the remaining letters are attached should be more rapidly solved, "on the average," than anagrams with many possible initial bigrams or trigrams (low ruleout).

While it is clear that the algorithmic process mentioned above is not carried out by *S*, it seems reasonable to assume that *S* may be performing a similar task on a more intuitive basis; consequently, the purpose of this experiment is to test the hypothesis that words which have a high ruleout total will be more quickly solved than those with a low ruleout total.

METHOD

Subjects.—Twenty graduate student volunteers from Indiana University served as *S*s. Twelve of the 20 were females.

TABLE 1

ANAGRAM SOLUTION TIMES (IN SEC.) FOR INDIVIDUAL ANAGRAMS AND FOR FOUR CONDITIONS OF WORD FREQUENCY AND RULEOUT

Cond.	Solution Word	Anagram	Mdn.	Range	Mdn. for Block	*M* for Block	*SD* for Block	Mean Anagrams Solved for Block
High Frequency, High Ruleout (90–94) *M* = 92	LIGHT MONTH STAND CHILD PLANT	ITLGH NMTHO TNDSA DHCIL LNTPA	4.0 6.5 17.5 6.5 12.5	1–36 1–180 5–180 1–147 1–180	9	32	16.9	18.2
High Frequency, Low Ruleout (72–76) *M* = 74	WATER LAUGH WHITE PRICE THEIR	WREAT HULAG HIETW ICREP RTHIE	100.0 28.0 45.0 131.0 17.0	19–180 4–180 5–180 2–180 2–180	44	77	22.5	15.0
Low Frequency, High Ruleout (90–94) *M* = 92	WRING FLINT SCANT CRAMP BLUNT	NWGRI NIFLT ANCST RAPCM BNTUL	20.0 8.0 180.0 25.0 50.0	5–180 1–180 2–180 3–180 2–180	26	68	23.1	15.0
Low Frequency, Low Ruleout (70–76) *M* = 74	CHAOS POUCH DEITY SCOUR BROIL	OCHSA HPCOU ITDYE OCURS OLBRI	180.0 22.0 180.0 71.5 36.5	10–180 2–180 25–180 4–180 3–180	69	99	24.0	12.0

Stimulus material.—The stimulus material consisted of 20 five-letter anagrams typed in capital letters on white 3 × 5 cards. The 20 words chosen for examination were selected from the Thorndike-Lorge (1944) word list at two frequency levels: 10 words from the first 500 words (occurring more than 100 times per million words) and 10 from the seventh thousand in frequency (occurring eight or nine times per million words).

Examination of "ruleout" totals for 75 five-letter words revealed a range from 66 to 94, with an *M* = 82 and an *SD* equal to 6.1. At each frequency level then, five words were chosen with "high ruleout" totals (90–94) and five with "low ruleout" totals (70–76).

All anagrams were presented in "hard" letter arrangements. That is, an attempt was made by the writer to present the anagrams in orders as dissimilar to the solution word as possible. Because of the interest in *initial* bigrams and trigrams, transition probabilities of solution words and anagrams were not controlled. The words and anagrams are presented in Table 1.

Procedure.—The anagrams were placed in a 20 × 20 Latin square to control sequence and practice effects. All *S*s were tested individually. Each was told that a five-letter anagram is five scrambled letters which, properly rearranged, will form one five-letter English word. They were also informed that they would have a maximum of 180 sec. to solve, and that if no solution occurred in that time, the correct solution would be provided before the next anagram was presented. No example was provided. Solutions were given verbally by *S* and times were recorded by means of a stopwatch.

RESULTS AND DISCUSSION

Available for analysis were 20 scores for each *S*. Because of the skewed distribution of the scores and the truncation imposed by the ceiling of 180 sec. per word, median time scores were selected as best representing each *S*'s performance. As a consequence, nonparametric statistical methods were used throughout the analyses.

As pointed out above, word frequency is a potent variable in predicting anagram solution time. Yet if the "ruleout" hypothesis has validity, high-frequency words with low ruleout totals should require considerable information processing before solution is reached; conversely, low-frequency words with high ruleout totals should require considerably less information processing and would consequently result in more rapid solution. As a result of this reasoning, *E* predicted the following order for the four groups of words used (arranged in order from fastest to slowest time to solution): (*a*) high frequency, high ruleout; (*b*) low frequency, high ruleout; (*c*) high frequency, low ruleout; (*d*) low frequency, low ruleout.

Page (1963) has developed a statistic (*L*) as a test of the null hypothesis against an ordered alternative. Thus, if the null hypothesis is expressed by $H_0: m_1 = m_2 = m_3 = m_4$, then the ordered alternative here is $H_1: m_1 < m_2 < m_3 < m_4$.

For purposes of this analysis the median of the scores for each block of five words for each *S* was obtained. These medians were then ranked, providing a matrix of rankings for the 20 *S*s. In the five cases of ties, a coin was tossed to assign the appropriate ranking.

The resulting *L* value (567) exceeds the critical value at the .001 level. Thus it may be inferred that significant agreement exists between the predicted order of anagram difficulty and the orders of solution times by *S*s.

The major concern of the present paper is the examination of the effects of high or low "ruleout" on anagram solution time. Figure 1 graphically illustrates the effects of "ruleout" on both high- and low-frequency words. Since the two lines are essentially parallel, it is possible to infer that no interaction occurs between frequency and "ruleout."

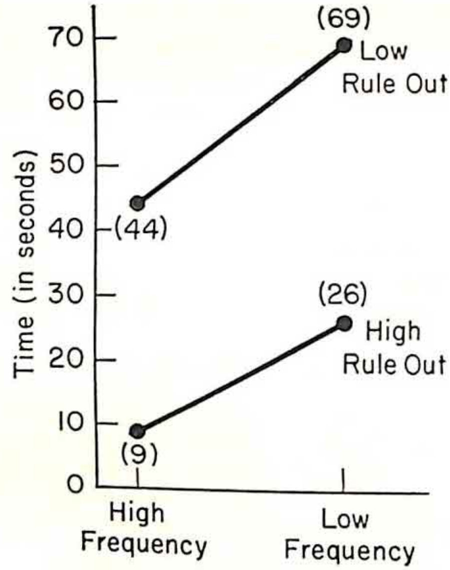


FIG. 1. Median time to solution for all words for all *S*s under each of the four conditions of frequency and "ruleout."

At each word frequency level the sign test (Siegel, 1956) may be used to test the following hypothesis:

$$H_0: P \geqslant \tfrac{1}{2}$$
$$H_1: P < \tfrac{1}{2}$$
$$N = 20$$

When

P = probabilty that $Md_{HRO} - Md_{LRO}$ is positive.

Where

Md_{HRO} = median solution time for high "ruleout" anagrams.

Md_{LRO} = median solution time for low "ruleout" anagrams.

For both high- and low-frequency words, anagrams with high ruleout were solved significantly more quickly than those with low ruleout ($p < .01$ in each case).

Five-letter anagrams constructed from high-frequency words are, as previous findings suggest, more quickly solved than those constructed from low-frequency words. However, the

present study suggests that for anagrams constructed from these low-frequency words, presence of relatively few initial bigrams which can logically be formed into English words when combined with the remaining letters results in more rapid solution than anagrams constructed from high-frequency words with many possible initial bigrams which will form English words.

REFERENCES

BEILIN, H., & HORN, R. Transition probability effects in anagram problem solving. *J. exp. Psychol.*, 1962, **63**, 514–518.

ERLEBACHER, A. *Parameters of anagram tasks.* (Doctoral dissertation, University of Wisconsin) Ann Arbor, Mich.: University Microfilms, 1962, No. 63–591.

HUNTER, I. M. L. The solving of five-letter anagram problems. *Brit. J. Psychol.*, 1959, **50**, 193–206.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56**, 376–379.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47**, 117–125.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of word transition probabilities. *J. exp. Psychol.*, 1962, **63**, 510–513.

MILLER, G. A. *Language and communication.* New York: McGraw-Hill, 1951.

PAGE, E. B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Amer. statist. J.*, 1963, **58**, 216–230.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received November 26, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 40–46

EFFECTS OF REPEATED BRIEF EXPOSURES ON THE GROWTH OF A PERCEPT [1]

RALPH NORMAN HABER [2] AND MAURICE HERSHENSON [3]

Yale University

Exposure duration and the number of times the word was flashed (trials) were varied independently in order to investigate the growth of the perception of a word. With duration constant, the probability of perceiving a word increased with exposure trials so that the word was quite clear and easily identified after a number of flashes, even if the first flash appeared blank. The function relating the probability of perceiving a word to the number of exposure trials could be specified knowing only the asymptote (maximum probability attainable) and the probability of perceiving the word on the first exposure. Despite this effect of repeated exposures, the probability of perceiving a word was always higher for a single flash at a given duration than for 2 or more flashes at shorter durations summing to the same total duration.

Word-recognition thresholds have been studied extensively in the past decade with particular emphasis on such variables as frequency, recency, word length, set, meaning, and value. Typically, the ascending method of limits has been used to determine thresholds defined as the duration at which a word is correctly recognized.

Little attention has been given to the detailed analysis of the perceptual processes involved in the growth of the perceptual experience of a word. From this point of view, the number of exposure trials becomes relevant as an independent variable and, along with exposure duration, more than an index of the effects of other variables.

Since exposure trials and exposure duration have been confounded in previous studies of word recognition using the method of limits, the present experiment was undertaken to study these variables independently. Specifically, this study was directed at assessing the effects of repeated exposures of words upon *S*'s ability to perceive them when exposure duration was invariant. Two questions guided the research: (*a*) Do repeated exposures of a word at a constant duration contribute to the growth of the phenomenal experience of the word? (*b*) If trials is a contributing variable, then in what way does it operate, and what is the relation between trials and the well-documented effects of exposure duration?

Moreover, with independent variation of trials and duration, it should be possible to assess the relative contributions of these variables to the development of the percept.

METHOD

Stimuli.—Each *S* was shown 504 seven-letter, three-syllable English words, selected from the Thorndike-Lorge (1944) word count and from the *Webster's Unabridged Dictionary* (1939). The most frequent two-thirds of the 615 seven-letter, three-syllable words from the Thorndike-Lorge list (general summary count—G) were used. An additional 500 words of the same structure were culled from Webster's. Judges, familiar with the differ-

[1] This research was partially supported by a grant from the National Institute of Mental Health (MH-03244) to the first author. We wish to thank Martha Breen for her assistance in the collection and analysis of the data and Michael Reed for his assistance in the mathematical analysis.

[2] Now at the University of Rochester.

[3] Now at the University of Wisconsin.

ences between rare and nonrare words in the Thorndike-Lorge list, selected 94 words that were judged to be nonrare. Thus, the 504 words represent a sample (virtually the population) of seven-letter, three-syllable words that appear not infrequently in print.

The words were divided into nine groups of 56 words each (corresponding to nine experimental sessions) according to a random process. They were printed on a 6-in. wide paper tape, using a LeRoy lettering stencil, No. 3240-290CL and pen, No. 3233-4. The letters were $\frac{3}{8}$ in. high and the words were $2\frac{1}{4}$ in. long. The words were drawn so as to equalize the spaces between the seven letters. The vertical spacing between the words was 1 in. The horizontal visual angle of the words was $2\frac{1}{2}°$.

The words were presented singly in the rear of one channel of a three-channel mirror tachistoscope (Scientific Prototype Manufacturing Corporation, Model D). A second, background, channel was always lighted, containing faint lines above and below the word which served as fixation boundaries. The stimulus durations were monitored before and after the experiment and varied by less than ± 0.1 msec. The reflectance measured at the eyepiece with a Macbeth illuminometer was 10 ft-L for the background and 18 ft-L for the stimulus plus background.

Procedure.—Each word was assigned to one of seven duration values, ranging from 5 to 35 msec. in steps of 5 msec., and one of eight exposure numbers (1, 2, 3, 4, 5, 10, 15, 25), representing the number of times it would be exposed. Thus, for each session, one word was assigned to each of the 56 Duration $\times$ Trial combinations. The assignment of words to these combinations was random for each session.

Since no prior threshold measurements had been made for *S*s, durations that were either too high or too low (in terms of lack of variability) were dropped after the first day. Thereafter, five duration values were used, although not necessarily the same ones for each *S*. Therefore, comparison between *S*s required that they be assigned relative values. The lowest of these five durations at which *S* perceived more than just an occasional word was defined as his threshold (T) duration. The other durations were designated as T $-$ 5, T $+$ 5, T $+$ 10, T $+$ 15. Approximately 12 words in each of the 5 $\times$ 8, Duration $\times$ Trial combinations were shown over the 9 days.

The stimuli were exposed when *S* pressed a button to trigger the T-scope. Hence, *S* was always prepared for the flash. He always knew when a word was being changed, but he never knew before an exposure whether that would be the last time he would see the word. The *S* was told that he would not know this and that he should treat each exposure as if it were the last one. Intertrial interval was never less than 8 sec.

For each stimulus exposure, *S* was required to report verbally the letters that he was certain he perceived and their positions, even when he was certain of the word. This was given great stress in the instruction. All of the analyses reported are based on the identification of letters, not of words. The *S* was said to have perceived the word if he correctly identified all of the seven letters on any of the exposures. This criterion, using letters rather than words, maximizes the probability of reports based on what *S* saw, rather than on what he thought he saw, or what he was sure must have been presented, but did not actually see.

Subjects.—Ten Yale undergraduates, enrolled in the introductory psychology course, served 9 hr., 1 hr. per day, over a period of 2 wk. They were tested individually.

Results

The data to be presented below can best be understood by first noting the phenomenal reports of *S*s as they participated in the experiment. If the duration of a word was low, *S* was usually unaware of letters or parts of letters on the first flash—the flash was blank. On the second or third flash, with no change in duration, beginnings of letters and sometimes whole letters would appear. After several more flashes, a number of letters would be present—often the whole word.

The percept of the word that developed after repetition was in no sense fuzzy, hazy, or the product of a guess. It assumed very clear status, so that *S* was never uncertain about his report, even though he was unable to see anything a few exposures earlier. These reports were a dramatic demonstration that perceptual thresholds must have been a function of repetition as well as energy and duration of stimulation.

TABLE 1

PERCENTAGE OF WORDS IN WHICH ALL LETTERS WERE CORRECTLY IDENTIFIED, AS A FUNCTION OF EXPOSURE DURATION AND NUMBER OF EXPOSURE TRIALS

Duration	Trials								Average
	1	2	3	4	5	10	15	25	
T − 5	0.0	0.9	0.0	1.0	0.0	1.9	1.9	2.9	1.1
T	7.3	19.2	25.6	38.7	33.9	51.7	58.2	62.3	37.1
T + 5	38.7	71.3	74.6	88.6	84.3	93.3	92.6	96.7	80.6
T + 10	61.9	89.0	92.5	90.6	96.7	98.4	98.3	99.2	91.7
T + 15	80.9	93.9	97.5	100.0	96.4	97.5	98.8	100.0	95.5

Note.—Ten *Ss* per cell; an average of 12 different words in each cell.

The data supporting these phenomenal impressions are given in Table 1 which presents the percentages of words perceived (i.e., all letters correctly identified) as a function of the duration at which the word was presented and the number of times it was exposed. It is clear that varying the number of exposure trials increased the likelihood of perceiving words.

Two functions, based on different models, produced reasonably close fits to the data points for three durations: T, T + 5, T + 10. One function is given by the expansion of the binomial $(p_1 + q_1)^n$, where p_1 is the probability of perceiving a word on Trial 1, $q_1 = 1 - p_1$, and n is the number of trials. This model is based on the assumption that the individual trial probabilities are independent of the number of trials, i.e., the probability of perceiving a word would be the same on Trial $n + 1$ as on Trial n. Further, this model predicts that trials should have a cumulative effect on accuracy such that as S is given more trials, he should be able to perceive a greater proportion of words. Thus, while the probability of perceiving a word *on* a given trial is constant, the probability of perceiving a word *by* Trial n is given by

$$P_n = 1 - q_1^n \qquad [1]$$

which increases with n. This function is plotted in Fig. 1 in broken lines.

The second function (fitted by the method of least squares, and plotted in Fig. 1 as solid lines) is

$$P_n = A - B/n^x, \qquad [2]$$

where A is the maximum probability of perceiving a word at a given duration (estimated on the basis of 25 trials), B is a constant, the meaning of which will be discussed below, and n is the number of trials. For the effects of trials on accuracy to be significant, the exponent of n (x) can not be zero. This exponent (respectively, 1.05, 0.85, and 1.14) differed significantly from zero ($p < .01$) for each of the three durations.

The asymptote of Function 2, or the maximum probability of a word being perceived for a given duration (in above terms), is given by

$$A = \lim_{n \to \infty} P_n,$$

and may take any value between zero and one depending upon the duration. However, the data indicate that A will be 1.00 except when the duration of the stimuli is very short.

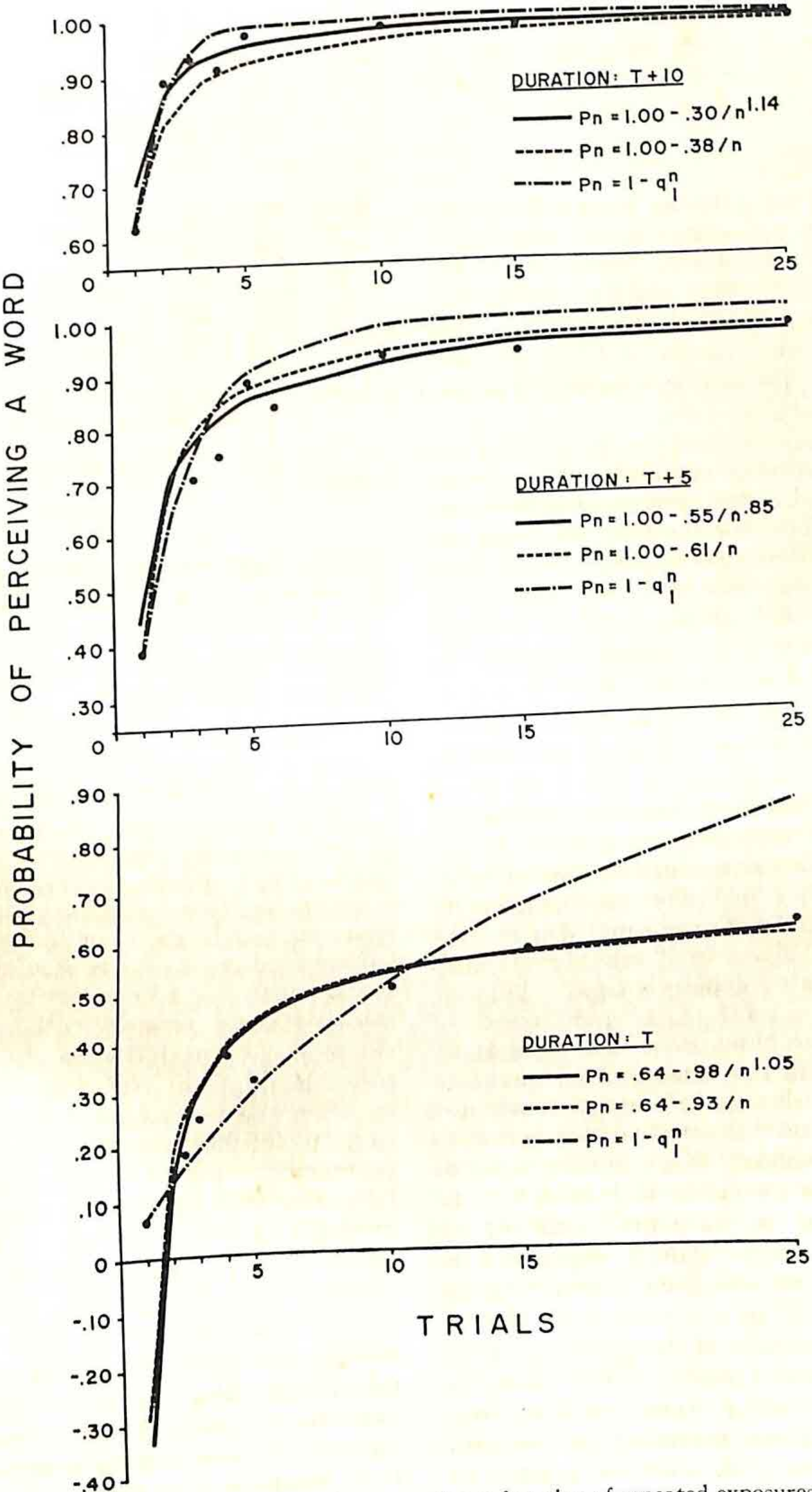


FIG. 1. Probability of perceiving a word as a function of repeated exposures.

The value of B was close to that of q_1 for all three durations and the exponents of n did not differ significantly from 1.00. The function

$$P_n = A - q_1/n \qquad [3]$$

is plotted in dashed lines in Fig. 1 and closely approximates the empirically derived functions. Thus, for all but relatively short durations where $A \neq 1.00$, the value of P_n can be calculated from estimates of q_1 obtained from a list of words presented for one exposure per word.

Within each of the three durations, the binomial expansion is generally below the data points (and the other functions) for the first few trials and above them thereafter. This implies that, for the first few trials, the probability of perceiving a word *on* a given trial is increasing from trial to trial. The converse seems to occur after about 10 trials, indicating that if *S*s have not perceived the word by then, the chances of their doing so on any given trial decreases with additional trials.

The relative contributions of trials and duration may be assessed by determining the optimal distribution of any given total number of milliseconds of exposure time. That is, given a total of, say, 50 msec. of exposure time, does one flash of 50 msec., or two flashes of 25 msec., or five flashes of 10 msec. maximally facilitate the veridical perception of the stimulus? Note that introducing trials as a variable while keeping total stimulus exposure time constant not only requires shorter exposures per trial as the number of trials increases, but also introduces a time interval between successive exposures. Table 2 illustrates that one long flash was always better than two or more shorter ones summing to the same total duration. This suggests that

TABLE 2

PERCENTAGE OF WORDS PERCEIVED WHEN THE TOTAL NUMBER OF MILLISECONDS OF DURATION IS DIVIDED INTO ONE OR MORE EXPOSURES

Total Msec.	Flash Duration				
	10	15	20	25	30
20	3.1		42.0		
30	5.1	26.1			74.1
40	9.2		67.5		
50	6.0			85.1	
60	7.0	37.7	68.3		93.7
75		35.6		91.4	
80	8.0		86.8		
90	8.5	37.0			97.3
100	9.1		81.7	91.2	
120	9.5	42.0	83.0		89.7
150	10.1	47.4		95.5	96.3

Note.—The number of flashes in each cell is obtained by dividing the total number of milliseconds by the flash duration. Trial No. 6, 7, 8, 9, and 12 were approximated using Function 3.

exposure duration was a more important variable than the number of exposure trials.

The relative contributions of duration and trials may be further analyzed by comparing the data of this experiment with threshold measurements obtained with more traditional methods where duration and exposure trials are confounded. For example, comparison can be made between the achieved accuracies after, say, the fourth flash at 25 msec., when either the first three flashes were also at 25 msec. in duration (this experiment), or when the first three flashes were, say, 10, 15, 20 msec. (traditional experiments). Figure 2 presents data from two such experiments: the new method curve is derived from results of the present experiment; the old method refers to the data collected on a comparable group of 10 *S*s using exactly the same procedures, equipment, and values as in the present experiment, except that the durations increased 5 msec. with every trial. It is clear that except for the first

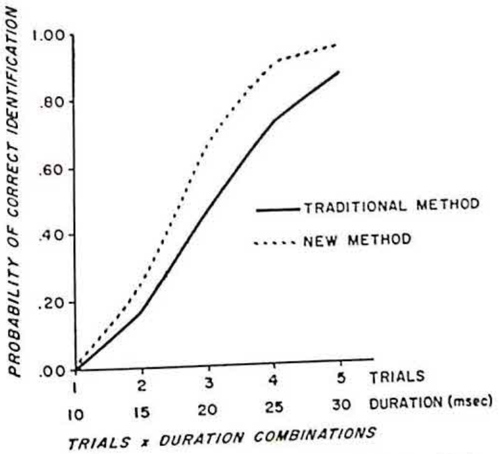


FIG. 2. Probability of correctly identifying a word as a function of method of combining trials and durations.

point, where, of course, there should be no difference, the new method results in higher accuracy. Thus, holding trials constant, the higher the duration of the earlier trials, the greater the accuracy.

Discussion

The data demonstrate that repeated exposures of a word at a constant duration does contribute to the growth of the phenomenal experience of the word. The prediction of accuracy achieved after any number of trials requires knowledge only of the probability of perceiving a word on the first trial, and the maximum probability after many trials. This information can be used in one of two different functions to predict accuracy: the binomial expansion based on the assumption that the individual trial probabilities are independent and constant from trial to trial, and the function $P_n = A - B/n$, where A is the maximum probability, B is the probability on the first trial, and n is the number of trials.

The second function seems to fit the data points better. It differs from the binomial in that it falls above the binomial for the first few trials, and below it for the remaining ones. Since the binomial is an index of independence, it appears that information received and processed on one flash facilitates perception of a word on the next flash for the first few trials. Thereafter, the effect appears to decay, and possibly becomes inhibitory in nature. This could be caused by S relaxing his effort after repeated failure to perceive the word, or comparing the information obtained with impossible or very unlikely words.

The latter reasoning is consistent with Bruner's (1957) early thinking about perceptual recognition. However, it is directly contrary to Bruner's more recent work (Bruner, 1963), where he finds that stimuli first presented in an impoverished form (Bruner currently uses a procedure where the stimuli are out of focus) are more difficult to recognize. Bruner suggests that this occurs because S is attempting to fit the stimuli to a hypothesis that is most likely wrong (because the stimulus was impoverished) and therefore takes extra trials to disconfirm that hypothesis before he can go on to check more correct ones. Applied to the present experiment, this reasoning would predict that the early trials, which are clearly impoverished, should fall below the binomial. That is, having been exposed to one impoverished flash, subsequent flashes should be harder to perceive. The data are contrary to this prediction. It is the early trials that are above, or at least equal to, the binomial, not below it. Further, this is most clearly the case for the threshold duration, where the stimuli were most impoverished.

A related problem concerns the failure of the threshold curve to approach 1.00. One explanation might be that the curve would approach unity if enough trials had been given. This appears unlikely from the shape of the curve.

A second alternative is that certain rare or difficult words were never identified when presented at threshold durations. However, an examination of the words at threshold duration indicated no tendency for specific words to be missed more frequently than other words. Moreover, since the words were ordered randomly with respect to duration, trials,

and days, it is unlikely that specific word characteristics could have systematically affected the results.

A third alternative warrants more detailed consideration. If the image of the stimulus from which *S* is working fades rapidly at very low durations so that *S* can not "scan" all of the letters, he may not be able to perceive them all before they fade. This interpretation is consistent with recent evidence (Averbach & Sperling, 1961) which suggests that short-term memory fades quite rapidly and allows *S* only a short time, perhaps much less than a second, to process the letters in the stimulus. If processing of a seven-letter word takes a given length of time, and the duration of the flash leads to a short-term memory shorter than that time, then *S* will never correctly identify all of the letters. This interpretation suggests that the threshold duration in this experiment was probably very near the threshold for short-term memory, so that, with oscillation, some of the words were retained long enough to be identified while others were not. In examining the raw protocols, many words were found for which *S*s saw all of the letters, but never on the same trial. This occurred a number of times, though it does not account for all the words which were never fully perceived. There were many words for which *S* never saw all of the letters even on the different trials.

If the short-term memory explanation of the failure to reach asymptote at unity is correct, then the effects of duration on perception can be reinterpreted. Duration would now be seen as a discontinuous function such that words presented below a given value (allowing for oscillation) would never be perceived regardless of the number of trials given. This is not to say that duration is no more than a dichotomous variable with respect to its effects on perception. There is ample documentation that increasing the stimulus duration will increase the accuracy of perception. However, this relationship leaves unanswered problems. For example, if the evidence regarding short-term memory is correct, why should variation of a few milliseconds in duration of the stimulus affect accuracy when presumably the stimulus produced a short-term memory lasting hundreds of milliseconds? Much more research is needed before the relation of duration of stimulation to perception becomes clear.

REFERENCES

Averbach, E., & Sperling, G. Short-term storage of information in vision. In C. Cherry (Ed.), *Information theory.* London: Butterworth, 1961.

Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

Bruner, J. S. Perceptual recognition as problem solving. Invited Address at Eastern Psychological Association, April 1963.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Webster's new international dictionary of the English language. (2nd ed., Unabridged) Springfield, Mass.: G. & C. Merriam, 1939.

(Received December 7, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 47–55

ASSOCIATION AND DISCRIMINATION IN PAIRED-ASSOCIATES LEARNING [1]

MARTHA C. POLSON, FRANK RESTLE, AND PETER G. POLSON

Indiana University

A paired-associates list with 8 unique items and 8 confusable twinned stimuli, was administered to 50 Ss by the anticipation method. Responses were 5 common nouns, highly integrated and discriminable, and prelearned by Ss. Unique items were learned in an all-or-nothing fashion, but twin items were learned in 2 steps, association and stimulus discrimination, each of which was all-or-nothing. A discontinuous multiple-process model, based on strategy-selection theory, was shown to fit the data.

One branch of the theory of paired-associates (PA) learning, stemming from E. J. Gibson's (1940) work on stimulus differentiation, says that a PA may be learned through several processes. Underwood and others (Underwood, Runquist, & Schulz, 1959; Underwood & Schulz, 1960) have extended the idea, and shown that along with stimulus discrimination there are separable processes of response discrimination and response integration. McGuire (1961) has partitioned errors into those caused by confusions between stimuli, those caused by lack of response integration, and associative failures. If *S* gives the response that belongs with a similar stimulus, then the error is attributable to stimulus confusion. If the response *S* makes is part of, or similar to, the correct answer, then the error is a failure of response integration or discrimination (Shepard, 1957). Omission or any other error is probably caused by lack of the correct association. This analysis of errors gives a picture of the processes involved in learning a PA list.

A second diverging branch of paired-associate theory begins with Rock (1957) followed by Estes (1959, 1960), Estes, Hopkins, and Crothers, (1960), and Bower (1961). These authors assert that ordinary PA learning occurs as an all-or-nothing process, and show that if stimuli are easily discriminated, responses are discrete and well known to *S*, and training proceeds at a leisurely pace, all-or-nothing (discontinuous) data may be observed.

Three main experimental methods have been used to test the all-or-nothing hypothesis; Rock's item-replacement method (Rock, 1957), Estes' miniature experiment (Estes, 1960; Estes et al., 1960), and the method of teaching a PA list by anticipation and then analyzing the data in detail (Bower, 1961; Bush & Mosteller, 1959). We employ the third method, using level of performance before the last error, the learning curve, $P(\mathcal{E}_n)$, the distribution of total errors (T), and the distribution of trial of last error (n'), as summary descriptions of the data.

[1] Preparation of this report was supported in part by National Science Foundation Grant No. GB 319 to the second author. The authors wish to thank Garry Flint and Gary Holstrom who helped in collecting, handling, and analyzing data; and David Kern for writing the program for generating the random orders and experimental procedures. The Indiana University Research Computing Center made a grant of machine time for these purposes, and it is gratefully acknowledged.

In Bower's all-or-none PA model S with respect to a given item begins in State S_0. His responses in S_0 are correct or wrong with probabilities p and q. On any trial with probability c he shifts to State S_1 on that item, and makes no more errors.

Before the last error, performance is stationary with a probability p of giving a correct response. The learning curve, probability of an error on Trial n, is given by

$$P(\mathcal{E}_n) = q(1-c)^{n-1} \qquad [1]$$

Distribution of Total Errors per Item (T_0):

$$P(T_0 = 0) = bp \qquad [2]$$

$$P(T_0 = k) = (1-bp)b(1-b)^{k-1} \quad \text{for any } k > 0.$$

Here b is the probability of no more errors following a given error, and is

$$b = \frac{c}{q+pc}$$

The distribution of total errors per item has mean

$$E(T_0) = q/c = u_0$$

and variance

$$\mathrm{Var}(T_0) = u_0 + (1-2c)u^2_0$$

Trial of Last Error (n'_0)

$$P(n'_0 = 0) = bp \qquad [3]$$

$$P(n'_0 = k) = qb(1-c)^{k-1} \quad \text{for all } k > 0.$$

The mean trial of the last error is

$$E(n'_0) = m_0 = bu_0/c$$

$$\mathrm{Var}(n'_0) = m_0\left(\frac{2}{c} - 1 - m_0\right).$$

An older all-or-nothing learning theory (called the "noncontinuity" or "insight" theory) has been applied to discrimination learning (Lashley, 1929). Several recent papers have revived this theory as a mathematical model (Bower & Trabasso, 1963; Restle, 1961, 1962; Trabasso, 1961, 1963). Assume that S begins with some strategy on which he bases responses. When a strategy fails, S resamples with replacement from his set of strategies. Suppose that strategies can be partitioned into irrelevant ones, which lead to success and failure with probabilities P and Q, and correct strategies which always lead to success. Let S_I be the presolution state of using an irrelevant strategy and S_L be the solution state of using a correct strategy. Let d be the proportion of correct strategies. The S will go from S_I to S_L only when his strategy fails, and then only with probability d. This model has the properties,

Learning Curve:

$$P(\mathcal{E}_n) = Q(1-Qd)^{n-1}(1-d), \qquad [4]$$

Distribution of Total Errors:

$$\begin{aligned} P(T_1 = k) &= (1-d)^k d, \\ E(T_1) &= (1-d)/d \\ \mathrm{Var}(T_1) &= (1-d)/d^2 \end{aligned} \qquad [5]$$

Distribution of Trial of Last Error:

$$P(n'_1 = 0) = d \qquad [6]$$

$$P(n'_1 = k) = (1-Qd)^{k-1}Qd(1-d) \quad \text{for all } k > 0.$$

$$E(n'_1) = (1-d)/Qd = m_1$$

$$\mathrm{Var}(n'_1) = m_1(2 + Qd - m_1).$$

The Bower association model and the noncontinuity theory of discrimination learning can be combined into a more comprehensive theory of PA learning. To obtain all-or-none PA data, Es eliminate all sources of difficulty except simple association. In particular, both stimulus discrimination and response integration are made very easy. However, PA experiments may involve several processes other than association. To account for the results of such ex-

periments we propose a *Discontinuous Multiprocess* theory in which each process is an all-or-nothing event. This idea has been developed and given detailed support by analysis of some existing data (Restle, 1963). A more definite test is made herein by first showing the distinctive mathematical properties of a discontinuous two-process model, then performing a detailed experimental test.

The model is as follows: Initially, in State S_0, S does not know the correct response and guesses among the response alternatives, having a constant probability p of being correct and distributing his errors among various responses. With probability c on any trial he learns the response to that item. If the stimulus is unique, S has mastered the item. If there is a confusable twin stimulus, then S may enter an intermediate State S_I in which he knows the response but cannot differentiate between stimuli. He therefore, with probability Q, makes a stimulus confusion error. On any error S resamples and may, with probability d, choose a discriminating strategy and stop confusing the twin stimuli. At that time he enters State S_L and since he has both formed the association and the discrimination, he performs correctly. Both steps of learning can occur on the same trial. The arrangement of states is as shown in Fig. 1.

The following statistical properties can be derived by elementary, though somewhat laborious methods:

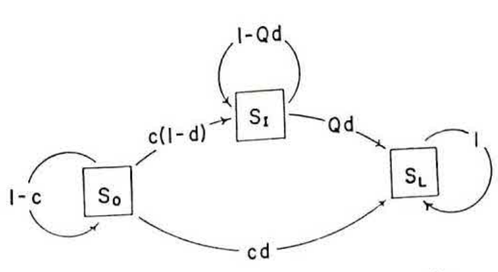


FIG. 1. Arrangement of states in the two-process model.

Learning Curve:

$$P(\mathcal{E}_n) = q(1-c)^{n-1} + \frac{Qc(1-d)}{c-dQ} \times [(1-dQ)^{n-1} - (1-c)^{n-1}] \quad [7]$$

Distribution of Total Errors:

$$P(T=k) = \frac{bd}{b-d} \times [(1-dp)(1-d)^n - (1-bp)(1-b)^n] \quad [8]$$

$$E(T) = E(T_0) + E(T_1)$$

from Equations 2 and 5.

$$\mathrm{Var}(T) = \mathrm{Var}(T_0) + \mathrm{Var}(T_1).$$

These results follow because total errors is the convolution (Feller, 1957) of the distributions of errors in States S_0 and S_I.

Trial of last error is dealt with in two parts. If no errors are made in State S_I (with probability d), then the last error is made in S_0, as given in Equation 3. If an error is made in S_I (probability $1-d$), then the trial of the last error is the convolution of the distribution of trials in State S_0 (not trial of last error in that state) with the distribution of trial of last error in S_I. This gives Trial of Last Error:

$$P(n'=k) = dbp \quad \text{for } k=0 \quad [9]$$

$$= dqb(1-c)^{k-1} + \frac{c(1-d)dQ}{dQ-c} [(1-c)^{k-1} - (1-dQ)^{k-1}] \quad \text{for } k>0.$$

$$E(n') = dm_0 + (1-d)\left[\frac{1}{c} + \frac{1}{dQ}\right]$$

For mathematical details and several related applications see Restle (1963).

This discontinuous two-process theory agrees in substance with McGuire's position that PA learning may involve separate processes, and that the processes can be separated experimentally and also by analysis of errors. The learning curves of the

separate types of errors are fit well by both theories.

The discontinuous theory is like McGuire's, but goes beyond. First, by defining specific states, it makes possible the use of ordinary discrete probability theory. Second, the discontinuous theory generates not only mean group effects, but distributions of statistics. An extra advantage of the discontinuous multiple-process theory is that it gives a definition of an irreducible process. The analysis of PA learning into separate processes is terminated whenever all processes yield geometric distributions. This prevents an infinite regress of subdivisions which might occur with continuous models and a determined theorist. To test the discrete two-stage theory, we performed a direct experimental analysis of the effect of stimulus similarity in PA learning, to see if it produces a separate and unitary stage of learning.

METHOD

A mixed-list design was used. Every *S* learned a list composed half of PAs with unique stimuli and half of PAs with confusable twinned stimuli.

Subjects.—In fulfillment of course requirements, 52 introductory psychology students served as *S*s. Two failed to learn in 30 trials, were found not to have understood instructions, and were discarded, leaving 50 *S*s in the analysis.

Apparatus and materials.—The *S*s sat at a table across from *E* and were shown flash cards over the top of an 18-in. screen. Stimuli were symbols and outline pictures; all 16 are shown in Fig. 2. As can be seen there were 8 unique and 8 twinned stimuli. Responses were the words COST, HOPE, PART, RUSH, and ONLY, which are among the most frequent 1,000 words (Thorndike & Lorge, 1944). Responses were assigned randomly to stimuli, separately for each *S*, with the restriction that twinned stimuli never both had the same response. Also no two pair of twinned stimuli had the same pair of responses for a given *S*.

Procedure.—Before beginning the PA task *S* was taught the five responses by having to repeat them in an order said by *E*. Five different orders were given until all were recited perfectly, and then *S* was required to repeat the words in any order. All *S*s succeeded.

In instructions, *S* was informed that some stimuli would be very similar to others; that each stimulus had one and only one correct response which would not change; and that each response would be the answer to more than one stimulus.

The stimulus card was shown until *S* responded, after which *E* gave the correct answer orally. Next was a 3-sec. delay while *E* recorded the response and selected the next

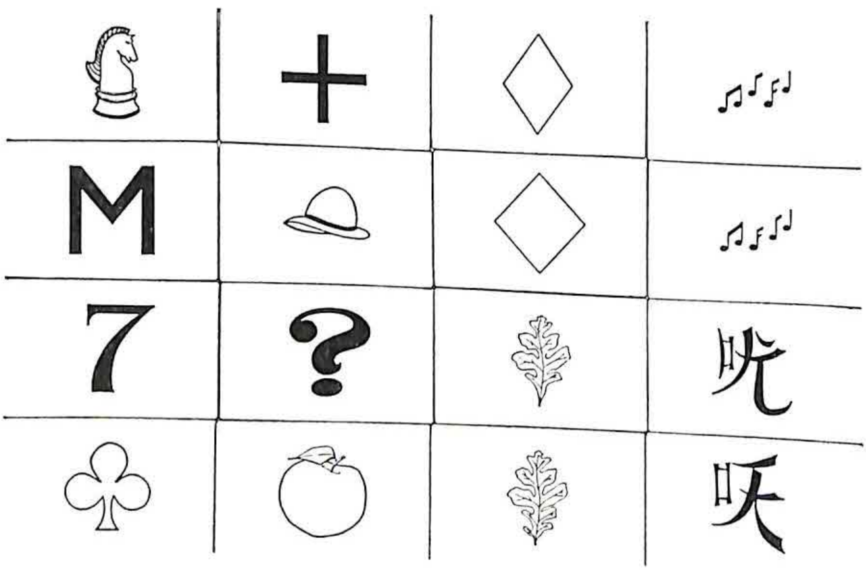


FIG. 2. Stimuli used in the experiment, showing twinned stimuli on the right.

stimulus. Training continued with no intertrial interval until *S* went twice through the list without error or completed 30 trials. Items were presented in a different permutation for each trial for each *S*, with the restrictions that twinned stimuli were not presented successively, and if a stimulus appeared in the fifteenth or sixteenth position on one trial it would not appear first or second on the following trial. Orders, subject to these restrictions, were generated randomly on an IBM 709 computer. The motivation behind most of these procedures is to produce all-or-nothing learning of the unique items. The use of short common words as responses, and slow flash-card training with face-to-face contact of *S* and *E*, were intended to replicate some of Bower's (1962) technique. The use of five responses for 16 stimuli, and the elaborate randomizing procedures, were intended to prevent efficient guessing. Five responses rather than, say, only two, were used so that confusions of twin stimuli could be distinguished from mere guesses.

Results

Items with twin confusable stimuli (twinned items) were much more difficult than items with unique stimuli (unique items). Mean errors were 5.96 and 2.80, respectively. Of the 5.96 errors, 3.81 were stimulus-confusion and 2.15 were nonconfusion errors.

Unique stimulus items.—These data closely approximate the all-or-nothing assumptions with $p = .30$, $c = .25$. Figure 3A shows the learning curve,

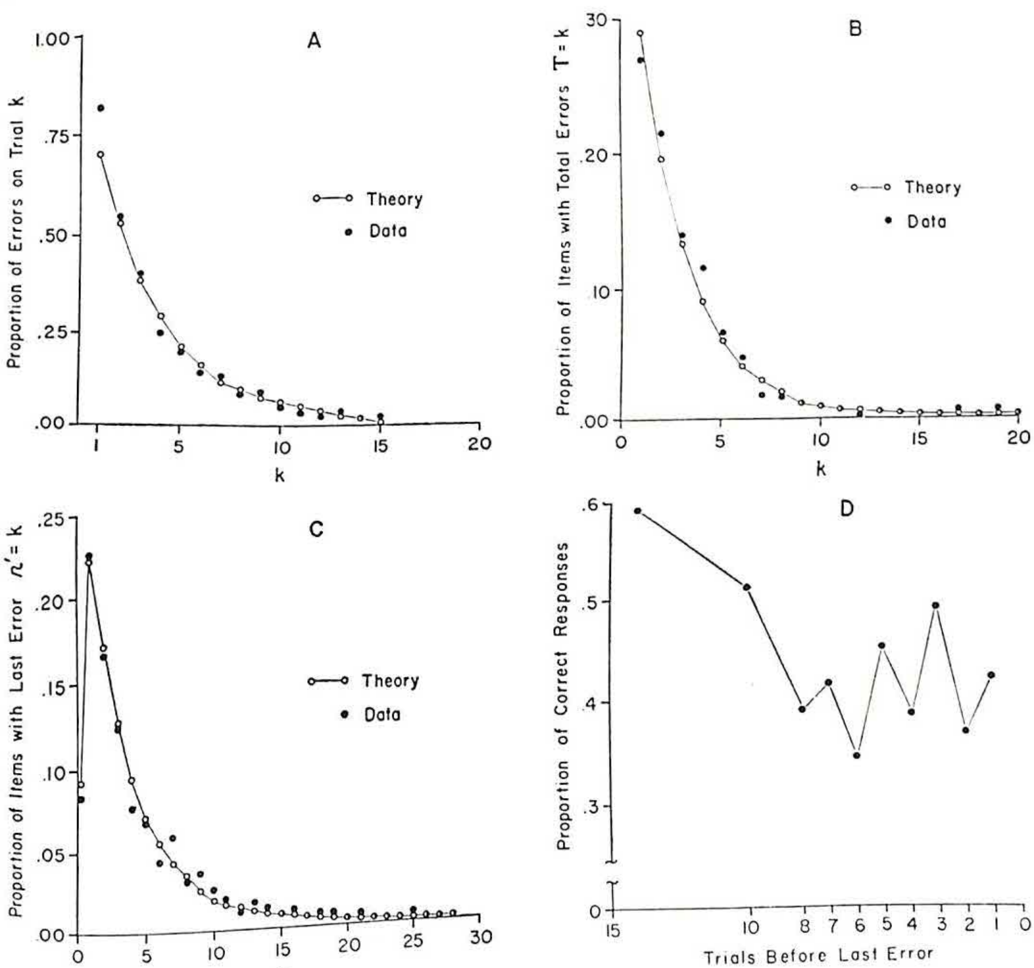


FIG. 3. Total errors on unique items: (A) as a function of trials (learning curve); (B) distribution of total errors; (C) distribution of trial of last error; (D) backwards learning curve.

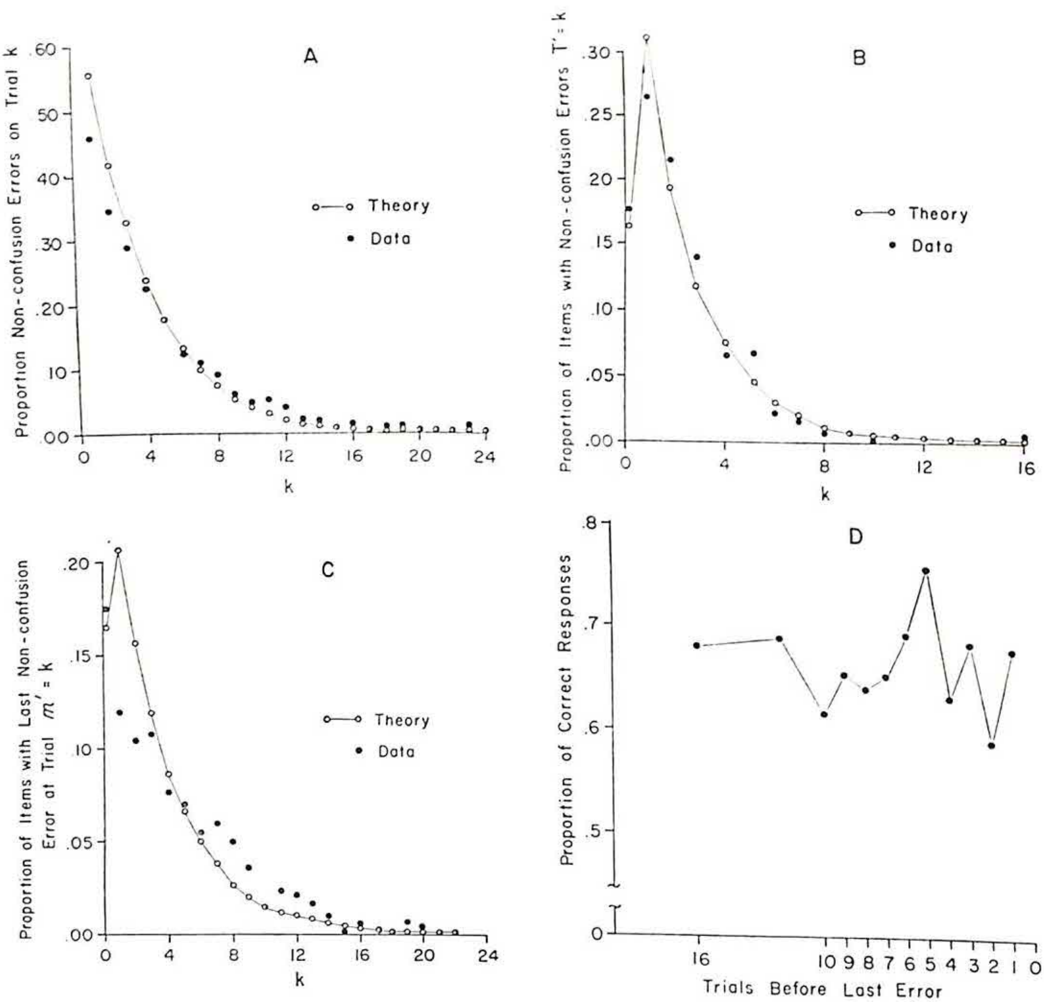


FIG. 4. Nonconfusion errors on twinned items: (A) as a function of trials (learning curve); (B) distribution of nonconfusion errors; (C) distribution of trial of last nonconfusion error; (D) backwards learning curve.

3B shows the distribution of total errors per item, 3C the distribution of trials of the last error per item. Theory and data are very close, comparable to Bower's (1961) results. There was some gradual improvement on trials before the last error, as shown by the backward learning curve in Fig. 3D.

Twin stimulus items.—Of the five responses available to *S*, one is correct, one is the confusion response, and the other three are scored as nonconfusion errors. Recall that on the average, *S*s responded erroneously with probability $1 - p = .70$ to the unique items which had four wrong answers. Since just three of these answers are scored as nonconfusion errors for twinned stimulus items, the probability of a nonconfusion error in State S_0 should be $.70 \times .75 = .525$. Theoretical predictions for $c = .25$ and $p = .475$, using the simple one-process model, were calculated. These are the appropriate parameters for twin items in State S_0, if we count only nonconfusion errors. Results are shown in Fig. 4, in which the four panels correspond to the four main statistical results. The predictions in Fig. 4 are made without any esti-

mated parameters; the value of c is from the unique items, and the predicted p depends upon a priori considerations and the results on unique items. Panel 4B, the distribution of total nonconfusion errors, shows an almost perfect correspondence. Panels 4A and 4C show that the errors are made on somewhat later trials than is predicted. The backwards learning curve (4D) shows stationarity but at a surprisingly high level, a mean of .625 correct, as contrasted with the calculated value of .475.

Now consider confusion and nonconfusion errors together. The mean total errors on twinned items is 5.96. Mean total errors on unique items was 2.80, and the difference, 3.16, is presumed to be the mean errors made in the intermediate state. By the strategy-selection theory (Restle, 1961) one can estimate the probability of learning, on an error, at $1/(1 + \bar{T})$ where $\bar{T}$ is mean errors. Taking account of the fact that some items were not mastered in 30 trials, $d \triangleq .225$. Since exactly two items are twinned in the list, it is assumed that the intermediate stage is one of two-choice discrimination. We set $Q = .50$. From

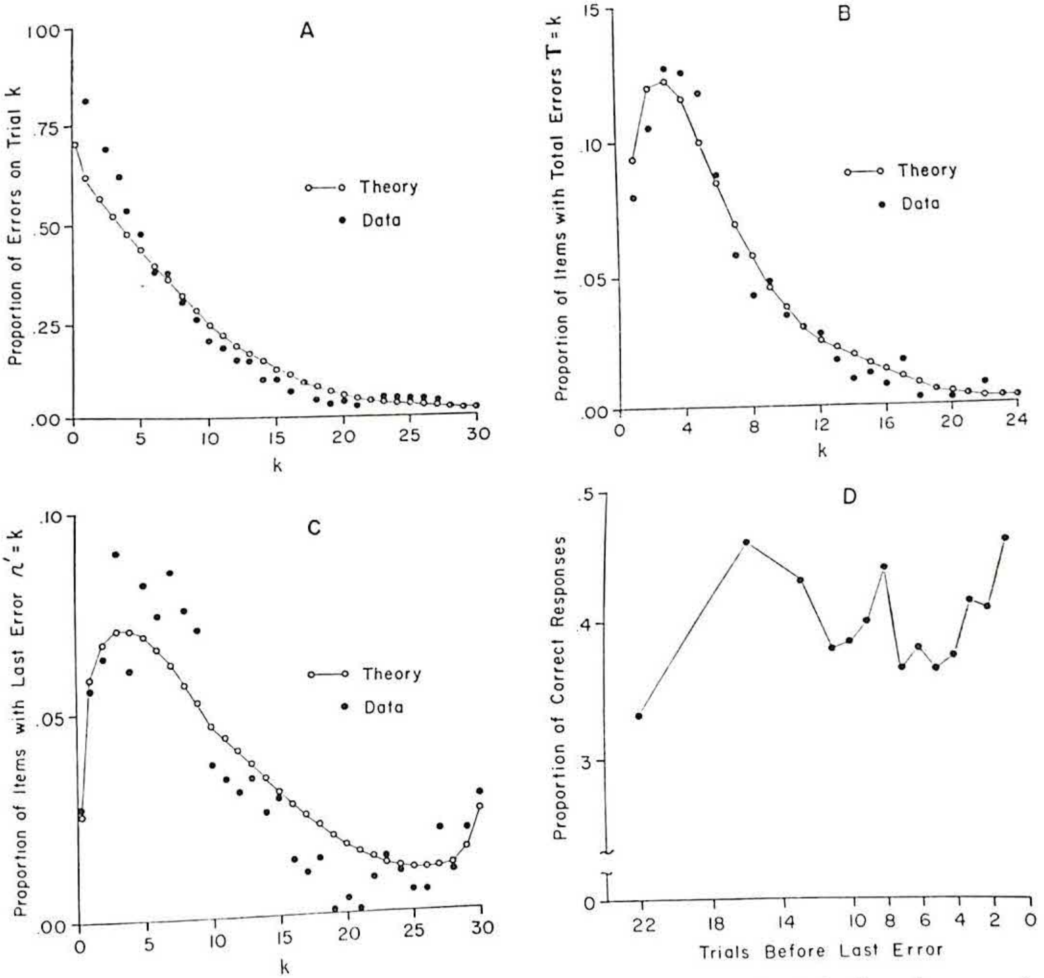


FIG. 5. Total errors on twinned items: (A) as a function of trials (learning curve); (B) distribution of total errors; (C) distribution of trial of last error; (D) backwards learning curve.

the unique items, $c \triangleq .25$, $p \triangleq .30$, and $b = c/(q + pc) = .33$. These values were inserted in Equations 7, 8, and 9 to compute predictions of total errors on twinned items. The calculations are compared with observations in Fig. 5. The theoretical curves in Fig. 5A, 5B, and 5C are calculated from Equations 7, 8, and 9.[2] The agreement is close, showing that total errors could arise from a two-stage process. A one-stage model will not fit these data.

Discussion

This experiment has demonstrated why there can be no answer to the question, "Is learning (especially PA learning) an all-or-nothing event?" In one list, with one set of *S*s, we obtained all-or-nothing data on half the items and two-process data on the other half. Each process is all-or-nothing and the times of the several processes are independent random variables. The several processes are identified as separate learning processes; in this we distinguish the present results from typical applications of Estes' "N-Pattern" theory (Estes, 1959).

Kintsch (1963) has recently shown that a two-process model fits the learning of PA items in which the responses require integration. He also shows the connection between these results and Rock's experimental method. Rock and Steinfeld (1963), using Rock's method, observed all-or-nothing performance when *S*s pronounced nonsense-syllable responses. Other *S*s who had to spell the responses, learned slower and showed multiple-stage learning, presumably because the spelled responses required integration and the pronounced words did not.

Above we have tested a simple additive model of the two processes of association learning and stimulus discrimination. In our model, each item is considered separately. A deeper analysis shows that twin stimuli are more intimately entangled than is said in that theory, and some of the consequences are seen in Fig. 4C and 4D. Figure 4C shows that the last nonconfusion (random) error was made on a later trial than predicted by the theory. What happens is that some items show initial guessing errors, then a string of confusion errors, then one or more guessing errors. The presolution curve, Fig. 4D, plots the proportion of correct responses on trials before the last random error. For this plot confusion errors are considered correct responses and only random errors are scored as errors. The blocks of confusion responses before a late random error have served to elevate this curve. It seems that items get into the stimulus discrimination state (State S_1) and then return to the original State S_0.

What may actually happen is this. Suppose that S-R_1 and S′-R_2 are two items and S and S′ are twinned. The *S* finds some strategy for remembering S-R_1 but then generalizes, and to S′ he says R_1. He knows no separate response to S′. Now he finds a discriminating strategy and no longer confuses S with S′. He makes R_1 to S, correctly, but he has no response at all to S′. He now makes random responses to S′ until he finds a strategy for remembering S′-R_2. Discrimination may occur before association on some items. Item-by-item inspection of protocols gives some support to this conjecture, in that if one item shows confusion followed by random errors, the twin item usually shows a string of correct responses throughout. This is consistent with the idea that one

[2] With the parameters estimated, we calculate that about 15 items would be left unsolved in the 30 trials, probably all in the intermediate discrimination State S_I. In Fig. 5B the total errors from Equation 8 are modified by adding in the distribution of 15 items unlearned for 30 trials. This is approximated by a binomial distribution of 30 random responses with $P = .5$. Figure 5C is modified by noting that of these 15 items, a fraction Q have last error at Trial 30, QP have last error at Trial 29, etc. This geometric distribution is added to the distribution calculated from Equation 9. These corrections for termination of the experiment are only approximate, for they do not take account of the possible effects of trials in State S_0.

name was learned for both stimuli, and then the stimuli were discriminated, before *S* ever learned the second name for the other stimulus.

A correct theoretical analysis requires at least that the pair of twin items be treated together, in a fairly complicated Markov chain analysis. We have not reduced such an analysis to manageable calculations thus far. Furthermore, a whole list of items may be learned by interconnected systems of strategies. The *Ss* may use sentences or other compound strategies to learn several items at once, or may classify stimuli, or use other techniques to master a list. Such strategies can be discovered by studying interconnections between learning of the various items in a list. We attempt no such analysis in this paper.

If the present theoretical position is correct, then (*a*) unitary learning is an all-or-nothing process, (*b*) many typical learning experiments require multiple processes to solution, and (*c*) these processes are each all-or-nothing, and with sufficient controls and proper experimental design can be separated.

REFERENCES

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255–280.

Bower, G. H. An association model for response and training variables in paired-associate learning. *Psychol. Rev.*, 1962, **69**, 34–53.

Bower, G. H., & Trabasso, T. R. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1963.

Bush, R. R., & Mosteller, F. A comparison of eight models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959.

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

Estes, W. K., Hopkins, B. L., & Crothers, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

Feller, W. *Introduction to probability theory and its applications.* (2nd ed.) New York: Wiley, 1957.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

Kintsch, W. All-or-none learning and the role of repetition in paired-associate learning. *Science*, 1963, **140**, 310–312.

Lashley, K. S. *Brain mechanisms and intelligence.* Chicago: Univer. Chicago Press, 1929.

McGuire, W. J. A multiprocess model for paired-associate learning. *J. exp. Psychol.*, 1961, **62**, 335–347.

Restle, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, **26**, 291–306.

Restle, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

Restle, F. Sources of difficulty in learning paired associates. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1963.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

Rock, I., & Steinfeld, G. Methodological questions in the study of one-trial learning. *Science*, 1963, **140**, 822–824.

Shepard, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–345.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Trabasso, T. R. The effect of stimulus emphasis on strategy selection in the acquisition and transfer of concepts. Unpublished doctoral dissertation, Michigan State University, 1961.

Trabasso, T. R. Stimulus emphasis and all-or-none learning in concept identification. *J. exp. Psychol.*, 1963, **65**, 398–406.

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired-associate lists as a function of intra-list similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received December 7, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 56–59

EXTINCTION AS A FUNCTION OF THE ORDER OF PARTIAL AND CONSISTENT REINFORCEMENT [1]

N. S. SUTHERLAND, N. J. MACKINTOSH, AND J. B. WOLFE

Oxford University, England

40 rats were trained in a straightway under partial or consistent reinforcement (PR or CR) or with one type of reinforcement schedule followed by the other, and were then extinguished. There was no significant difference in resistance to extinction between Ss given PR training alone, and Ss given CR after PR; but giving CR before PR significantly reduced resistance to extinction. These results are in disagreement with both the discrimination hypothesis and a cognitive dissonance account of PR, but confirm a prediction made by a new explanation of the PR effect.

Theios (1962) using rats in a straightway, found that resistance to extinction was almost as great when consistent reinforcement (CR) training was given after partial reinforcement (PR) as when PR training alone was given. Jenkins (1962) confirmed this result with pigeons, and also found that if birds were given 10 blocks of CR and 3 blocks of PR, resistance to extinction was greater if PR was given before than if it was given after CR. Since, however, 3 blocks of PR may not have been sufficient to produce the basic PR effect, this result is difficult to interpret. The aim of the present experiment is to test the relative effects of CR followed by PR and of PR followed by CR with sufficient PR trials to produce a clear-cut PR effect.

The results of such an experiment are of considerable theoretical interest. Sutherland (1964) has suggested that the PR effect may be understood in terms of a two-stage model of learning, in which it is assumed that animals must learn both which features of the stimulus situation to attend to (which analyzers to switch in), and which responses to make (Sutherland, 1959). The strength with which a given analyzer comes to be switched in depends upon the consistency with which its different outputs are differentially correlated with subsequent events of importance to the animal. Under CR training whichever analyzers are initially switched in will yield outputs which are consistently correlated with reward, and the running response will be conditioned only to the outputs from these analyzers. Under PR, however, no analyzer will yield an output consistently correlated with reward, and the animal should continually try out new analyzers: the running response will therefore be conditioned to the outputs from many analyzers. Provided enough trials are given for the strength of response attachments to be approaching asymptote, PR will produce more resistance to extinction than CR, since after PR training the response must be extinguished to the outputs from many more analyzers than after CR.

This model predicts that when an animal receives both PR and CR training, trials to extinction will largely be determined by the earlier reinforcement schedule. If PR is

[1] This work forms part of a project on "stimulus analyzing mechanisms"; the project is supported by the American Office of Naval Research (Contract N62558-2453) and the Nuffield Foundation.

given first, the response will be conditioned to the outputs from many analyzers, and subsequent CR training should not weaken these response connections. Giving CR training first will cause a small number of analyzers to be strongly switched in, and this will prevent the response being conditioned to the outputs from other analyzers under subsequent PR training. Thus the model predicts that giving CR after PR will lead to greater resistance to extinction than giving CR before PR. The opposite prediction appears to be made by at least two other explanations of the PR effect, namely, cognitive dissonance theory (Lawrence & Festinger, 1962) and the discrimination hypothesis (Bitterman, Fedderson, & Tyler, 1953); these theories predict that resistance to extinction will be at least as great when PR is given after CR as when it is given before.

Method

Subjects

The *S*s were 40 female hooded rats from the colony maintained at the Institute of Experimental Psychology, Oxford. They were experimentally naive, and aged 3 mo. at the start of the experiment.

Apparatus

The apparatus was a 6-ft. straight alleyway. The floor was made of chipboard and was 6 in. wide. The 6-in. high walls were made of hardboard. The first 7 in. was separated from the rest of the alley by a transparent Perspex guillotine door, and served as a start box. A similar guillotine door 3 in. before the end of the alley was used to prevent retracing. The goal box was 9 in. square, constructed of similar material, and always contained a circular feeding bowl, 2 in. high and 4 in. in diameter. The apparatus was unpainted, and covered with $\frac{1}{2}$-in. wire mesh.

Procedure

Pretraining.—A week before the start of the experiment all *S*s were put on a feeding schedule of 1 hr. food daily. Two days pretraining were given. On the first day, *S*s were placed in the empty apparatus in groups of four, and allowed to explore for 30 min. On the second day, they were each given two rewarded runs.

Training.—Twenty trials were given daily with a 6-min. intertrial interval (during which *S*s were kept in individual detention cages). Running times were recorded with a stopwatch to the nearest 0.2 sec. from the raising of the start-box door to the time *S*'s nose passed the guillotine door just before the goal box. On rewarded trials, *S* was allowed to eat in the goal box for 10 sec. On unrewarded trials *S* was detained in the goal box (with an empty feeding bowl) for 10 sec.

Extinction.—All groups started extinction on the day following their final acquisition trials. Twenty trials were given daily, with the same intertrial interval as during training. If *S* failed to reach the goal box in 2 min., it was removed from the apparatus and kept in its detention cage until its next trial. All *S*s were given 20 extinction trials on Day 1; thereafter they were stopped when they reached a criterion of three consecutive 2-min. trials.

Experimental Design

There were five groups each of 8 *S*s. All groups were treated alike during extinction, but they received the following different treatments during training. Group P were given 60 PR trials, and no CR trials; Group P-C received 60 PR trials followed by 100 CR trials; Group C-P received 100 CR trials followed by 60 PR trials; Groups C-60 and C-160 had no PR training and received, respectively, 60 and 160 CR trials.

Results

Table 1 shows for each group the average number of trials to the cri-

TABLE 1

TRIALS TO EXTINCTION

Group	Mean Trials to Extinction	*SD*
P	82	41
P-C	77	36
C-P	33	17
C-60	17	6
C-160	17	7

terion of extinction, namely, three successive failures to run in under 2 min. The differences between groups were evaluated by the Mann-Whitney U test; the following comparisons are of interest. (*a*) The Partial group extinguished much more slowly than the Consistent ($U = 0$, $p < .001$): hence, there was a clear-cut PR effect. (*b*) There was no significant difference in resistance to extinction between the two Consistent groups. (*c*) Comparing the Partial group with the two mixed groups, there was no significant difference between Groups P and P-C, but the difference between P and C-P was highly significant ($U = 5$, $p < .001$): thus giving CR before PR definitely reduces resistance to extinction as compared with PR training alone. (*d*) Both mixed groups are more resistant to extinction than either of the Consistent groups: the effect is only just significant for Group C-P ($U = 13$, $p < .05$), and is significant at the .01 level for Group P-C ($U = 9$): thus giving PR before CR markedly increases resistance to extinction as compared to giving CR training alone. (*e*) The most important comparison for our purposes is that between the two mixed groups: Group P-C was significantly slower to extinguish than Group C-P ($U = 7$, $p < .005$): thus the order in which partial and consistent reinforcement are given does affect resistance to extinction.

Running speeds are shown graphically in Fig. 1 for the last 10 trials of training and the first 60 trials of extinction (animals which met the extinction criterion were counted as having a running time of 2 min. on subsequent trials). Group P-C ran slower than Group P over the first 30 trials of extinction, but they continued running for almost as many trials. The difference between mean

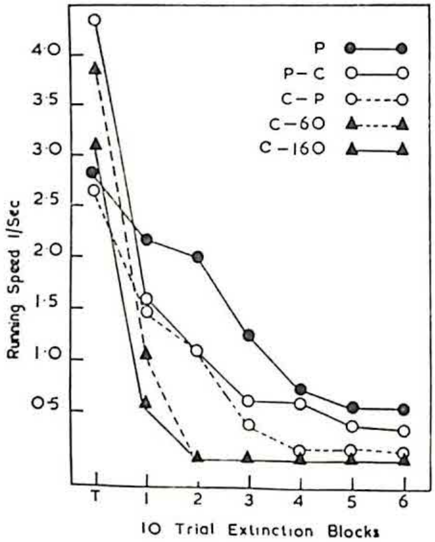


FIG. 1. Mean running speeds on last 10 trials of training (T) and first 60 trials of extinction.

running speeds of individual Ss in the two groups is significant at the .05 level of confidence ($U = 11$) over Extinction Trials 11–30.

DISCUSSION

The results confirm and extend those of Theios (1962) and Jenkins (1962). Both Es found (as we did) that giving CR training *after* PR led to only a small reduction in resistance to extinction. Our main result is that giving CR before PR produces a large reduction in resistance to extinction. This confirms Jenkins' finding, and demonstrates that the reduction occurs in a situation where there is a difference between the effects of PR training alone and CR training alone. The latter difference was particularly large in the present experiment, and this may be because only two trials pretraining with CR were given before training proper began: most Es give some pretraining with CR before PR (e.g., Theios gave 30 trials), and this should weaken the effect of PR and indeed may account for the failure of some investigators to obtain a clear PR effect.

The predictions made by three theories

of PR have already been discussed in the Introduction. Since giving CR before PR weakens resistance to extinction more than giving it after PR, the predictions made by the discrimination theory and by Lawrence and Festinger's theory are not confirmed. The prediction made by Sutherland's proposal is confirmed. Moreover, if CR strengthens some analyzers, the shape of the extinction curve for the P-C groups is explained. If the dominant analyzers stay switched in longer after CR training, the initial effects of extinction will be to weaken the responses attached to them (for a direct demonstration of this effect cf. Mackintosh, 1963b): hence, there will at first be a marked decrease in running speed. Only when the dominant analyzers have been switched off in the course of extinction will the effects of the initial PR training which attached the response to less dominant analyzers have a chance to show. Thus over the first 20 trials of extinction there is little difference in the running speeds of the P-C and C-P groups. A difference emerges only when the dominant analyzers are extinguished and at this point the C-P group extinguishes rapidly since the initial CR training has prevented the response from being attached to other analyzers, while the P-C group continues to run since the response is now controlled by the less dominant analyzers, and must be extinguished to them before running stops.

The present experiment is of course only indirect confirmation of the theory. Other results (in preparation) show that with PR animals learn about more aspects of the stimulus situation than with CR. Moreover, the PR effect is only one aspect of learning to which the model applies. Mackintosh (1962, 1963a, 1963b) has developed and tested a series of predictions made by the model concerning other aspects of the learning process.

REFERENCES

Bitterman, M. E., Fedderson, W. E., & Tyler, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

Jenkins, H. M. Resistance to extinction when partial reinforcement is followed by regular reinforcement. *J. exp. Psychol.*, 1962, **64**, 441–450.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcements.* Stanford: Stanford Univer. Press, 1962.

Mackintosh, N. J. The effects of overtraining on a reversal and nonreversal shift. *J. comp. physiol. Psychol.*, 1962, **55**, 555–559.

Mackintosh, N. J. The effect of irrelevant cues on reversal learning in the rat. *Brit. J. Psychol.*, 1963, **54**, 127–134. (a)

Mackintosh, N. J. Extinction of a discrimination habit as a function of overtraining. *J. comp. physiol. Psychol.*, 1963, **56**, 842–847. (b)

Sutherland, N. S. Stimulus analyzing mechanisms. In, *Proceedings of a symposium on the mechanisation of thought processes.* Vol. 2. London, England: Her Majesty's Stationery Office, 1959. Pp. 575–609.

Sutherland, N. S. Visual discrimination in animals. *Brit. med. Bull.*, 1964, **20**, 54–59.

Theios, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 1–6.

(Received December 2, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 60–66

EFFECTS OF DELAY INTERVALS AND MEANINGFULNESS ON VERBAL MEDIATING RESPONSES [1]

MARGARET JEAN PETERSON

Indiana University

Delay intervals of 0, 2, or 8 sec. were interpolated between the 2 acquisition stages of a mediation paradigm or between the second acquisition stage and the test stage. When the learning materials were of relatively low meaningfulness, mediated facilitation was found with delays of 0 and 2 sec. but not with a delay interval of 8 sec. When the learning materials were of high meaningfulness, mediated facilitation was observed with an 8-sec. delay interval as well. Response availability was inversely related to the length of the delay intervals.

The purpose of the studies was to assess the effects upon verbal mediation of delay intervals interpolated between the two acquisition stages of a mediation paradigm or between the second acquisition stage and the test trial. The importance of temporal contiguity upon the establishment of mediating associations has been emphasized since the days of Aristotle. Recently, using a simple reproduction task, Peterson, Peterson, and Miller (1961) demonstrated that forgetting may occur within seconds. If the factors producing forgetting operate during the acquisition of mediating associations, mediated responses would be expected to decrease as the length of the delay intervals was increased.

Two other variables known to influence verbal mediation were investigated simultaneously with the length of the delay intervals: the order of presentation of the items during the acquisition stages and the meaningfulness of the learning materials.

In addition to the response measure of mediated facilitation, a second measure, response availability, was also evaluated. Response availability was defined as the difference between empirical and a priori estimates of responding under a control condition.

METHOD

Experiment CCC₁.—Consonant trigrams, CCCs, of from 0 to 29% association value (Underwood & Schulz, 1960) were used to construct mediation sets. The 20 consonants were used repeatedly among the 648 letters comprising the trigrams used: the median frequency of repetition was 19. Repetitions of the letters of the trigrams making up a single mediation set were minimized as much as possible.

Table 1 presents the paradigms used to construct mediation sets (Column 2), sample mediation sets (Column 3), and sample control sets (Column 4) for two of the experiments. The order of presentation of the stages began with the top row and proceeded down the column, row by row. As illustrated in Table 1, Column 2, a mediation set began with the first stage presentation of a pair of consonant trigrams, A-B. Stimulus A was then shown alone as a cue trial and *S* was instructed to read off the stimulus and to give the response which had been paired with it on the previous presentation. The first delay interval, DI_1, followed. In the second acquisition stage, the response term from the first pair, B, was shown as the stimulus with another trigram response term, C, which was then followed by the appropriate cue trial, B, and the second delay interval, DI_2. In the test stage, the first stimulus trigram, A, was presented with the second stage response, C, and two additional response

[1] This research was supported by Grants M-5209 and 07944 from the United States Public Health Service, National Institutes of Health.

TABLE 1

SAMPLES OF THE LEARNING MATERIALS FOR TWO EXPERIMENTS

Exp.	Paradigm	Sample Sets	
		Mediation	Control
Exp. CCC_I	A-B A DI_1 B-C B DI_2 A-CDE	QJH XZF QJH (Name colors) XZF MWB XZF (Name colors) QJH MWB FHJ GXM	QJH XZF QJH (Name colors) ZGW MWB ZGW (Name colors) QJH MWB FHJ GXM
Exp. KR_{VIII}	(B-C)[a] (DI_1)[b] B-A B DI_2 A-CDE	(heavy light)[a] (?)[b] heavy GIL heavy (Name colors) GIL fast soap light	(poor—?)[a] (?)[b] poor GIL poor (Name colors) GIL fast soap light

[a] Assumed to have been established preexperimentally.
[b] Delay interval greater than 8 sec. filled in unknown fashion.

alternatives, D and E. The stimulus A was printed above the row of three response alternatives. The *Ss* were instructed to verbally pair each of the response trigrams successively with the stimulus trigram, A, and then to select the response alternative (i.e., C, D, or E) which made the "best pair" with Stimulus A. A control set was prepared for each mediation set by substituting a new extraneous trigram for Stimulus B in the second-stage pair, X-C, providing a pool of 54 mediation and their respective 54 control sets. Table 1 shows a sample mediation set and its control set in Columns 3 and 4.

The C trigram appeared equally often in the first, second, or third position among the response alternatives on the test trials. To equate frequency of experience with each of the test alternatives, the D and E trigrams for a particular set were trigrams used as the C trigram in other sets.

Cue trials were inserted after each acquisition stage to test the retention of the previous pair before proceeding with the rest of the set. If *S* erred on a cue trial, the preceding pair was reshown, then the cue trial was presented again, etc., until *S* responded correctly on the cue trial.

The lengths of the first delay interval, DI_1, (0, 2, or 8 sec.) were assigned factorially with the lengths of the second delay interval, DI_2, (0, 2, or 8 sec.). Thus, the experimental design consisted of a $2 \times 3 \times 3$ repeated measures factorial in which type of presentation (mediation-control), the first delay interval (0, 2, 8 sec.), and the second delay interval (0, 2, 8 sec.) were mutually orthogonal. The resulting 18 cells were each represented three times among the 54 sets assigned to a particular *S*. The order of presentation of the sets so assigned was randomized.

Experiment CVC_I.—Consonant-vowel-consonant trigrams, CVCs, of 100% association value (Underwood & Schulz, 1960) were used as the learning materials in Exp. CVC_I. The median frequency of repetition of the 648 letters used to construct the sets was 25. In all other respects, Exp. CVC_I was an exact replica of Exp. CCC_I.

Experiment CVC_{VIII}.—The CVCs of Exp. CVC_I were used as the learning materials for Exp. CVC $_{VIII}$ but the order of presentation of the items during the acquisition stages differed. The paradigm used in Exp. CCC_I and CVC_I (A-B, B-C, A-CDE) corresponded to Paradigm I of Peterson, Colavita, Sheahan, and Blattner (1964). They tested eight combinations of the acquisition stage pairs and reported the greatest evidence of mediated facilitation with Paradigm VIII (B-C, B-A, A-CDE). Paradigm VIII was used in constructing the mediation and their control sets for Exp. CVC $_{VIII}$. (The subscripts for the experiments identify the paradigm designation.)

The A trigram from the first mediation set of Exp. CVC_I was used as the A trigram for the first mediation set of Exp. CVC_{VIII}. Similarly, the B, C, D, and E trigrams from a particular set of Exp. CVC_I were transposed to their corresponding position in the Para-

digm VIII format. All other details were identical to Exp. CCC_I and CVC_I.

Experiment KR_{VIII}.—The CVCs of Exp. CVC_I and CVC_{VIII} became the A components for Exp. KR_{VIII}, while the B components were stimulus words from the Minnesota revision of the Kent-Rosanoff norms (Russell & Jenkins, 1954) as shown in Table 1. The C components were the most frequently occurring responses (35–83%) to the chosen stimulus words. The median frequency of repetition of letters used in constructing the sets was 20. Paradigm VIII provided the format for the 36 sets.

Assuming that the B-C stage had been established prior to experimentation, the paradigm investigated in Exp. KR_{VIII} was actually B-C, preexperimentally established, then B-A, and the test trial, A-CDE. Thirty-six mediation sets and their 36 control sets were prepared—sample sets appear in Table 1.

Only the second delay interval could be experimentally manipulated and the same lengths were used: 0, 2, or 8 sec. The first delay interval was assumed to exceed 8 sec. in all cases. The experimental design of Exp. KR_{VIII} was a 2×3 repeated measures factorial with type of presentation (mediation-control) orthogonal to the second delay interval (0, 2, 8 sec.) The six cells were each represented six times among the 36 sets assigned to an *S*. The order of presentation of the sets was randomized.

Apparatus.—All learning materials were printed on 4×6 in. index cards. The delay intervals were filled with a color-naming task. Three samples of each of 10 different colors were randomized and mounted on a memory drum tape. The colors were shown at a 1-sec. rate. Two Hunter interval timers controlled the lengths of the delay intervals and the rotation of the memory drum.

Experiment-subject scheduling.—To facilitate experimentation, the four experiments were conducted separately. The order in which the experiments were run was randomized, yielding the sequence Exp. CVC_{VIII}, CCC_I, KR_{VIII}, and CVC_I. All four were completed within the first 2 mo. of the spring semester, 1963. The 132 introductory psychology students at Indiana University who signed up for participation were assigned randomly to the experiments subject to the restriction that 36 *S*s be assigned to Exp. CCC_I, to CVC_I, and to CVC_{VIII} and 24 *S*s to Exp. KR_{VIII}. The *S*s were notified when to report for participation.

Procedure.—During the acquisition stage, *S*s read aloud the letters of the syllables or read off the words. On the cue trials, they were instructed to respond with the missing member of the pair after reading aloud the first member of the pair. If *S* erred on the cue trial, the pair was presented again, then the cue trial was shown, etc., until *S* responded correctly. The timers and the memory drum were activated when *S* said the last (sixth) letter of the response on cue trials. During the test trials, *S* successively paired each of the three response alternatives with the stimulus and then selected the response which made the "best pair" with the stimulus. No temporal restrictions obtained *within* the stages: however, *S*s tended to respond rapidly.

Three practice sets using different components in all positions were shown. The delay intervals were 2 and 8 sec. for DI_1 and DI_2 of the first practice set; 0 and 2 sec. for the second practice set; 8 and 0 sec. for the third.

RESULTS

Experiments CCC_I, CVC_I, and CVC_{VIII}.—Experiments CCC_I, CVC_I, and CVC_{VIII} were analyzed in one analysis of variance since *S*s assigned to the experiments were essentially random samples from the same population. Experiment KR_{VIII} was analyzed separately because only the second delay interval was varied. The mean number of C selections on the test trials for Exp. CCC_I, CVC_I, and CVC_{VIII} did not differ significantly nor were any interactions involving the three experiments reliable.

Mediated facilitation was shown by the reliably greater mean number of test-trial selections of C following all mediation presentations taken over the delay intervals (1.72) than following all control presentations (1.66), $F(1, 105) = 10.81$, $p < .01$. The interaction of the type of presentation with the experiments was not statistically significant.

The variable of primary interest, the length of the delay intervals interpolated, was associated with a reliable decrease in the number of selections of C on the test trials as the intervals were lengthened from 0 to 8 sec.

The mean numbers of Test-Trial C selections corresponding to delay intervals of 0, 2, and 8 sec. separating the two acquisition stages were 1.85, 1.75, and 1.60, $F(2, 210) = 11.64$, $p < .01$. The means for the 0-, 2-, and 8-sec. delay intervals occurring between the second acquisition stage and the test stage were 1.79, 1.73, and 1.67, $F(2,210) = 3.76$, $p < .05$. The means for the two delay intervals following mediation and following control presentation appear in Table 2. Although the interaction with the type of presentation was not statistically significant for either the first, $F(2, 210) = 1.38$, or the second, $F(2, 210) = 1.24$, delay interval, inspection of Table 2 suggested that the difference between selections of C on test trials following mediation and following control presentation (mediated facilitation) did tend to decrease as the length of the delay intervals was increased. A series of t tests conducted between the mean number of Test-Trial C selections following mediation presentation and following control presentation for each delay interval separately showed that mediated facilitation was obtained with 0- and 2-sec. delays but not with an 8-sec. delay for both positions of interpolation of the delay interval.

The other response measure investigated, response availability, was defined as the difference between the mean number of test-trial selections of C following control presentation with the a priori chance level of one selection of C per three test trials. It was assumed that an increase in the likelihood of selecting C on the test trials reflected some kind of sensitizing effect from the presentation of C during the acquisition stages. Response availability declined reliably over the delay intervals for both positions of interpolation of delays.

Since the enhancement of the test appearance of C was assumed to result from the recency of its acquisition-stage presentation, the paradigm with C in the first acquisition stage (VIII) should show a lower index of response availability than the paradigm with C presented during the second acquisition stage (I). The mean for Exp. CVC_{VIII} was 1.26 and the mean for Exp. CVC_{I} (restricting the comparison to identical learning materials) was 1.59, supporting earlier findings that response availability was related to the distance between the acquisi-

TABLE 2

Mean Number of Selections of C Alternative on Test Trials

	Delay Interval 1			Delay Interval 2		
	0	2	8	0	2	8
Mediation presentation CCC_{I}, CVC_{I}, CVC_{VIII}	1.95	1.80	1.63	1.86	1.83	1.69
Control presentation CCC_{I}, CVC_{I}, CVC_{VIII}	1.72	1.70	1.57	1.72	1.64	1.64
Mediation presentation KR_{VIII}				2.36	2.42	1.90
Control presentation KR_{VIII}				.86	1.23	1.02

tion presentation of C and its appearance as a test-stage alternative (Peterson et al., 1964). Following the same line of reasoning, it would be expected that the interpolation of delay intervals between the two acquisition stages would reduce the index of response availability to a greater extent for Exp. CVC_{VIII} with the C component presented during the first acquisition stage than for Exp. CVC_{I} with the C component presented during the second acquisition stage. This contention was also supported: the decrease in the response availability index from a 0-sec. delay to an 8-sec. delay for DI_1 with Exp. CVC_{VIII} was .26 and the comparable decrease for Exp. CVC_{I} was .06. The decrease in the response availability index for DI_2 was greater for Exp. CVC_{I}, .14, than for Exp. CVC_{VIII}, .00. These results indicated that the reliable decline in response availability associated with DI_1 was produced primarily by Exp. CVC_{VIII} and the decline associated with DI_2 was produced primarily by Exp. CVC_{I} and CCC_1 (.12).

Experiment KR_{VIII}.—The mean number of test-trial selections of C following mediation presentation averaged over the delay intervals was 2.23 and following control presentations, was 1.04, $F\ (1, 23) = 49.11$. This difference was reliable beyond the .01 level, again demonstrating mediated facilitation. The mean numbers of test-trial selections of C following second delay interval values of 0, 2, and 8 sec. were 1.61, 1.82, and 1.46, $F\ (2, 46) = 7.42$. These differences were reliable.

As suggested by the means in Table 2, the interaction between Type of Presentation and the Delay Interval was statistically significant, $F\ (2, 46) = 5.26$, $p < .01$. Further analyses showed reliable evidence of mediated facilitation for all three delay intervals, $t_{\text{Med.-Control at 0 sec.}}\ (46) = 11.03$; $t_{\text{Med.-Control at 2 sec.}}\ (46) = 8.71$; $t_{\text{Med.-Control at 8 sec.}}\ (46) = 6.43$. When the mean numbers of selections of the C alternative on test trials following mediation presentation associated with the various delay intervals were analyzed, the difference between 2 and 8 sec. was reliable, $t\ (46) = 3.82$, $p < .01$ but the difference between 0 and 2 sec. was not, $t\ (46) = .44$, $p > .05$. The reverse was true for the mean number of selections of the C alternative following control presentation: only the difference between 0 and 2 sec. was reliable, $t\ (46) = 2.76$, $p < .01$. As with Exp. CVC_{VIII}, the second delay interval appeared to influence mediated facilitation to a greater extent than it influenced response availability.

With Exp. KR_{VIII} the C component was not actually experienced during the acquisition stage, rather it was assumed to have occurred during the preexperimental history of *S*. It is possible that response availability increased with longer delay intervals because *S*s had more time to think of associations to the B components (Kent-Rosanoff stimulus words) that were presented during the acquisition stage.

Discussion

As the length of the delay intervals increased, the estimates of mediated facilitation decreased, particularly when learning materials of relatively low meaningfulness were used. When the length of either delay interval was 8 sec., no evidence of mediated facilitation was found for the experiments using CCCs and CVCs. When common English words were incorporated in the mediation paradigm, Exp. KR_{VIII}, reliable evidence of facilitation was found with a delay interval of 8 sec. although it was lower than the estimates associated with 0- and 2-sec. delay intervals. Previously, Peter-

son and Blattner (1963) reported that mediated facilitation increased as the meaningfulness of the learning materials increased. However, their observations would not explain the greater susceptibility to delay intervals of the less meaningful materials. Some insight may be provided by a study which showed that retention increased with increases in the meaningfulness of the stimulus materials (Peterson et al., 1961). In the present situation, it was expected that the more meaningful materials would be more likely to be retained until the test trials than were less meaningful materials. The prediction assumed that some representational form of the acquisition associations had to be retained until the test trials if mediated facilitation were to be demonstrated.

No test for retention was introduced at the *end* of each delay interval because the temporal separation of the stages was one of the variables under investigation so no direct evidence of retention *following* the delay interval was obtained. Each association was tested for retention at the *beginning* of each delay interval to be sure the association could be repeated by *S*. The mean number of errors made on the retention or cue trials of Exp. CCC_I was 13.75; of Exp. CVC_I and CVC_{VIII} were .83 and .83; and of Exp. KR_{VIII} was .00. One interpretation was that the lower the meaningfulness, the more errors made on cue trials, and the more susceptible the estimates of mediated facilitation were to interference from interpolated delay intervals. As Underwood and Schulz (1960, p. 4) have suggested, the variable of meaningfulness appears to wield a great influence over verbal measures, including mediation. Another interpretation of the results was suggested by the error analysis. The acquisition pairs of Exp. CCC_I, in particular, may not have been learned as well as the acquisition pairs of Exp. KR_{VIII}, and these differences might have been reflected in the estimates of mediated facilitation obtained on the test trials. However, two features of the experimental procedure should be noted. Whenever errors were made on the cue trials, the pair and the cue trial were repeated until *S* responded correctly on the cue trial so that the pairs of Exp. CCC_I were actually presented more often than the pairs of the other experiments. Because *S*s were required to read off the stimulus material whenever it was presented, at least a certain amount of familiarity, if not learning, should have accrued to the pairs of Exp. CCC_I. In addition, when the cue-trial responses were correct, the pairs were not shown again, thus denying the experimental opportunity for overlearning which might create differences in the degree of learning between experiments with overlearned and with barely learned pairs.

Turning to the paradigms used during the acquisition stages, Exp. CVC_{VIII} was expected to show greater evidence of mediated facilitation than was Exp. CVC_I because Paradigm VIII yielded more mediated facilitation than any other paradigm in an earlier study (Peterson et al., 1964). No reliable differences were observed when all sets were considered in the comparison or when the comparisons were restricted to an analysis of the means of sets with no errors on the cue trials, although the predicted trend was found in the latter analysis.

The interpolation of delay intervals influenced response availability also. The greatest effect occurred when the delay interval was introduced immediately after presentation of the C component. Thus for Exp. CVC_{VIII} with C presented as the first-stage response term, the decrease in response availability was more marked than the decrease observed when the delay interval followed the second acquisition stage and the reverse was true for the two experiments with B-C learned as a second-stage association (CCC_I and CVC_I). The effect of interpolation of a delay interval immediately after the acquisition presentation of B-C was not investigated with Exp. KR_{VIII} because B-C was assumed to have been established prior to actual experimentation but more than 8 sec. must have intervened.

REFERENCES

PETERSON, L. R., PETERSON, M. J., & MILLER, A. G. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

PETERSON, M. J., & BLATTNER, K. C. Development of a verbal mediator. *J. exp. Psychol.*, 1963, **66**, 72–77.

PETERSON, M. J., COLAVITA, F. B., SHEAHAN, D. B., III, & BLATTNER, K. C. Verbal mediating chains and response availability as a function of the acquisition paradigm. *J. verbal Learn. verbal Behav.*, 1964, **3**, 11–18.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216, Office of Naval Research.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received December 11, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 67-74

DIFFERENTIAL CONDITIONING, EXTINCTION, AND SECONDARY REINFORCEMENT

ROGER W. BLACK[1]

University of Tennessee

The present study was concerned with secondary reinforcement effects in extinction. 82 rats were given differential conditioning in a pair of straight alleys. For each S one alley (black or white) was positive (a sucrose solution reward in goal box) while the other was negative. Following 42 acquisition trials, Ss were divided into 4 groups. G++ was extinguished in the previously positive alley and G−− in the negative alley. G−+ was extinguished in an alley consisting of the positive runway (RW) and negative goal box (GB), while for G−+ the RW was negative and the GB positive. Starting, running, and goal-box speeds were recorded. G−+ showed the slowest rate of extinction, while extinction was generally fastest for G+−. G++ and G−− were intermediate. These results were interpreted as contrary to the conventional concept of secondary reinforcement and an alternative formulation was suggested.

In 1953 Bitterman, Feddersen, and Tyler reported that (*a*) resistance to extinction following partial reinforcement of a running-jumping response is lessened if the goal box (GB) is altered at the beginning of extinction; (*b*) following discrimination training, the previously negative (unrewarded) GB produces greater resistance to extinction than the positive GB; and (*c*), that *S*s which receive discrimination training and are then extinguished with the previously positive GB show less resistance to extinction than *S*s which had received ordinary partial reinforcement during acquisition.

While Finding *a* can readily be interpreted as a secondary reinforcement effect, the second result seems quite embarrassing to that concept, in that *S*s receiving more secondary reinforcement (i.e., exposure to the positive GB) during extinction, extinguished faster than *S*s receiving less such reinforcement. Elam, Tyler, and Bitterman (1954), however, reported a subsequent replication of this result.

One feature of the Bitterman and Elam studies which complicates interpretation is the fact that they measured response strength at a point in the alley at which no differential, reward-related cue was present. Thus, although GBs of different brightnesses were used on positive and negative trials, the same gray alley was employed on all trials. While this procedure may make the secondary reinforcement properties of the GBs obvious, the reinforcement properties of the alley—where response time was actually recorded—are ambiguous.

The present study makes possible a more definitive analysis of the role of secondary reinforcement in extinction. Four groups of *S*s were given differential conditioning training in a pair of straight alleys, a runway (RW) and GB of one brightness being positive and the other negative. The groups were then extinguished under the following conditions: for Group ++ the positive RW and GB were em-

[1] The writer expresses his appreciation to Gordon Kimbrell and Robert Fowler for their aid in the collection of these data.

Now at the State University of Iowa.

oyed; for Group − −, the RW and B were negative. For Group + − ne RW was positive, while the GB as negative and for Group − + the egative RW and positive GB were sed. This design permitted the comarison of groups receiving different B cues (or degrees of secondary einforcement) but the same RW cues, s well as groups receiving different W cues, but the same GB cues. In ct, two such comparisons under ach condition were possible. Thus, ne comparisons of Group + − vs. roup − − and of Group + + vs. roup − + should each reflect secondry reinforcement effects in the RW ith GB cues constant; while the omparison of Group + + vs. Group − and of Group − + vs. Group − should reflect secondary reinorcement in the GB with constant W cues. According to the concept f secondary reinforcement, of course, ne first member of each comparison hould show better performance in xtinction than the second, since in ach case the first member receives ne greater amount of secondary reinrcement—i.e., exposure to a stimus previously associated with priary reward.

Method

Subjects.—The *S*s were 90 male albino rats rchased from the Budd Mountain Rodent rm, Chester, New Jersey and were about days old at the beginning of deprivation.

Apparatus.—The apparatus consisted of a r of enclosed straight alleys 3 in. wide and n. high with a start box (SB) 6 in. long, 4-in. runway (RW), and a 9-in. goal box B).

A metal start door separated the SB and /, while between the RW and GB was a od door hinged at the top of the alley ch *S* opened to enter the GB and which vented *S* from seeing the GB from the . The interior of one RW and GB was ted a flat white, while the other RW and were black. Both SBs were gray. The s were attached to the RWs by hooks and thus could be interchanged. A metal-lined indentation at the center of a wood block served as the food cup in each GB.

By means of floor treadle switches and Standard Electric timers, starting, running, and GB times were recorded. Start time was measured from opening of the start door to 6 in. within the runway; running time over the next 12 in.; and GB time over the next 12 in., and, thus, to a point 6 in. within the GB.

Procedure

Drive maintenance and pretraining.—One week prior to the beginning of pretraining, all *S*s were placed on a 20-hr. food deprivation schedule with water continuously available. Four days prior to the beginning of acquisition, a pretraining schedule was initiated in which *S* was placed in a walnut-stained box 12 × 8 × 6 in. high, containing a glass caster into which had been placed 1 cc of 32% by weight sucrose solution. The *S* remained in the box for 5 min. or until it consumed the sucrose solution. Five such feedings were given each day for 4 days. Eight *S*s consistently failed to consume the sucrose and were discarded.

Acquisition.—Before acquisition the animals were divided into two groups of 41 *S*s each. For one group, the white RW and GB were designated as "positive" while for the other group, the black RW and GB were "positive." On each day of acquisition *S* was placed into one of the SBs, the start door was opened and *S* remained in the alley until it entered the GB. A 1-cc 32% sucrose solution served as reward on positive trials, no solution being present on negative trials.

On the first three trials of the first day of acquisition the door between the RW and GB was open and on the following three trials it was half open. Six trials were given each day with an average intertrial interval of about 15 min. On the first day all trials were in the positive alley, while on the remaining 6 days of acquisition, three trials were in the positive, and three in the negative alley, the order of trials (positive vs. negative) being randomized separately for each day.

Extinction.—On the eighth day following the beginning of acquisition, extinction began. The *S*s were divided into four subgroups defined in terms of the RW-GB combination in which they were extinguished (the same gray SB used in acquisition was used for all groups in extinction). Group + + ($n = 21$) was tested during extinction in an alley consisting of the RW and GB which had been positive for those *S*s during acquisition; for

Group $+-$ ($n = 20$) the positive RW but negative GB was employed; for Group $-+$ ($n = 20$) the negative RW and positive GB; and for Group $--$ ($n = 21$) both the RW and GB were negative. In both Group $+-$ and Group $-+$, white was positive for 10 *S*s and black was positive for 10. In both Group $++$ and Group $--$, there were 10 white- and 11 black-positive *S*s.

Six trials per day for 7 consecutive days were administered during extinction. If *S* failed to enter the GB within 3 min. after an extinction trial began, it was removed from the alley and latency for that trial was recorded as 3 min.

RESULTS

Before analysis of the data was undertaken, all starting, running, and goal-box times were converted into speed scores (ft/sec).

Acquisition data.—Figure 1 indicates daily mean starting speeds to the positive and negative alleys during acquisition. (Times were not recorded on the first day of acquisition.) Inspection indicates that *S*s began to show clear differentiation between the alleys by the fifth day of training and by the final day of training mean starting in the positive alley was about twice that in the negative. The acquisition data for running and GB speeds were essentially identical to that for starting speed and are not presented.

Extinction data.—Because the speeds for the various groups converged with continuing extinction trials, only the data for the first 18 trials (3 days) were analyzed. When separate analyses of variance were performed on the three sets of speed scores, the relationships among the groups were identical for starting and running speeds, although somewhat different on GB speed. Therefore, only starting and GB speeds are discussed below.

Goal-box speeds.—Figure 2 represents level of performance during extinction in terms of GB speeds. Inspection indicates considerable differences in the performance of the four groups, during the early extinction. Thus Group $+-$ and Group $--$ show less response strength than Group $-+$ and Group $++$. Group

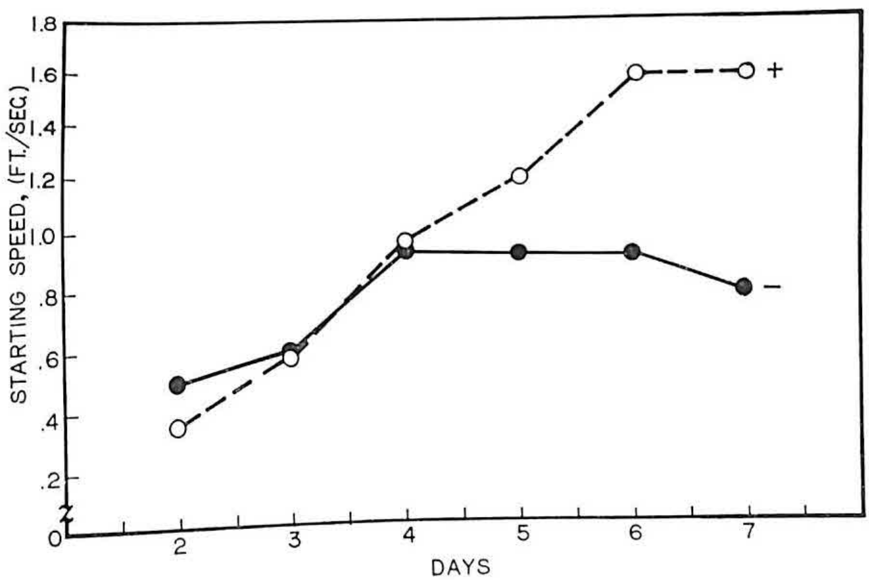


FIG. 1. Mean starting speed in the positive and negative alleys as a function of acquisition day.

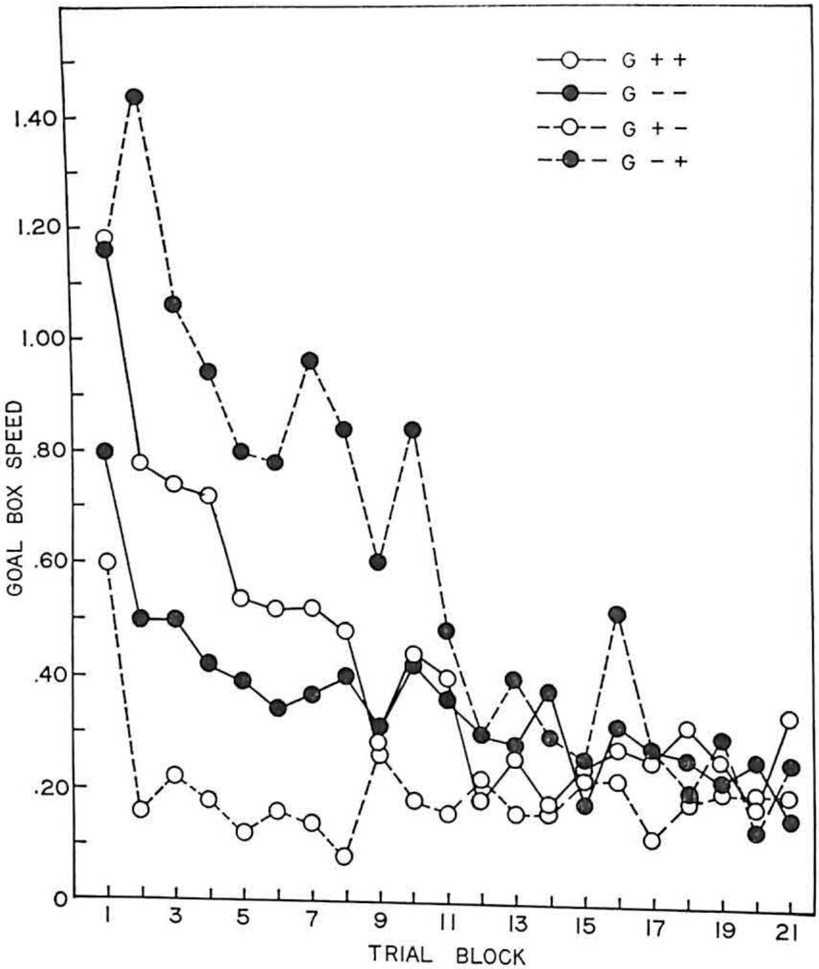


FIG. 2. Mean GB speed during extinction as a function of blocks of two trials.

+ −, in fact, appeared to reach virtually complete extinction in only 2 trials, while Group − + improved during the first 4 trials. Meanwhile Group + + was superior to Group − −, although their curves appeared to completely converge in about 18 trials. An analysis of variance indicated differences between groups on Trials 1–18 to be highly reliable, $F(3, 78) = 19.07, p < .001$, while the interaction of groups by trials during this period failed to prove significant, $F(51, 1326) < 1.00$.

While the preceding indicates *level of performance* during extinction, it should be noted that the four groups entered extinction at different levels of performance and hence a comparison of the *rate* at which the groups extinguished from their respective initial levels is of interest. Following a procedure described by Anderson (1963, p. 164) estimates were obtained for each *S*'s initial GB speed, R(1), and asymptotic extinction speed, R(0). R(1) was defined as an *S*'s GB speed on Extinction Trial 1, while R(0) was *S*'s mean speed on the last day of extinction. Using the expression:

$$\frac{R(0) - (1/N')\Sigma R(n)}{R(0) - R(1)}$$

where R(n) is S's GB speed on any trial from 1 to 18 and $N = 18$, a *rate of extinction score* was computed for each S. Means of these scores for the groups were: Group $+-$, .049; Group $--$, .408; Group $++$, .479; Group $-+$, .997. Thus, in agreement with the earlier analysis, the average GB speed of Group $+-$ for Trials 1–18 was only slightly above its final extinction level. Group $-+$, on the other hand, showed an average GB speed during these trials almost equal to its first trial speed of .84 ft/sec. An analysis indicated differences in extinction rate to be highly significant, $F(3, 78) = 13.32$, $p < .001$. Subsequent t tests indicated that Group $-+$ extinguished more slowly than any other group (all p's $< .01$) while Group $+-$ extinguished more rapidly than any other group (all p's $< .01$). The difference in rate of extinction between Group $++$ and Group $--$, however, failed to prove reliable, $t(40) = 0.64$.

Starting speeds.—Inspection of Fig. 3 indicates substantial group differences in performance during extinction and an analysis indicated these differences were highly significant, $F(3, 78) = 7.39$, $p < .001$, while the Groups

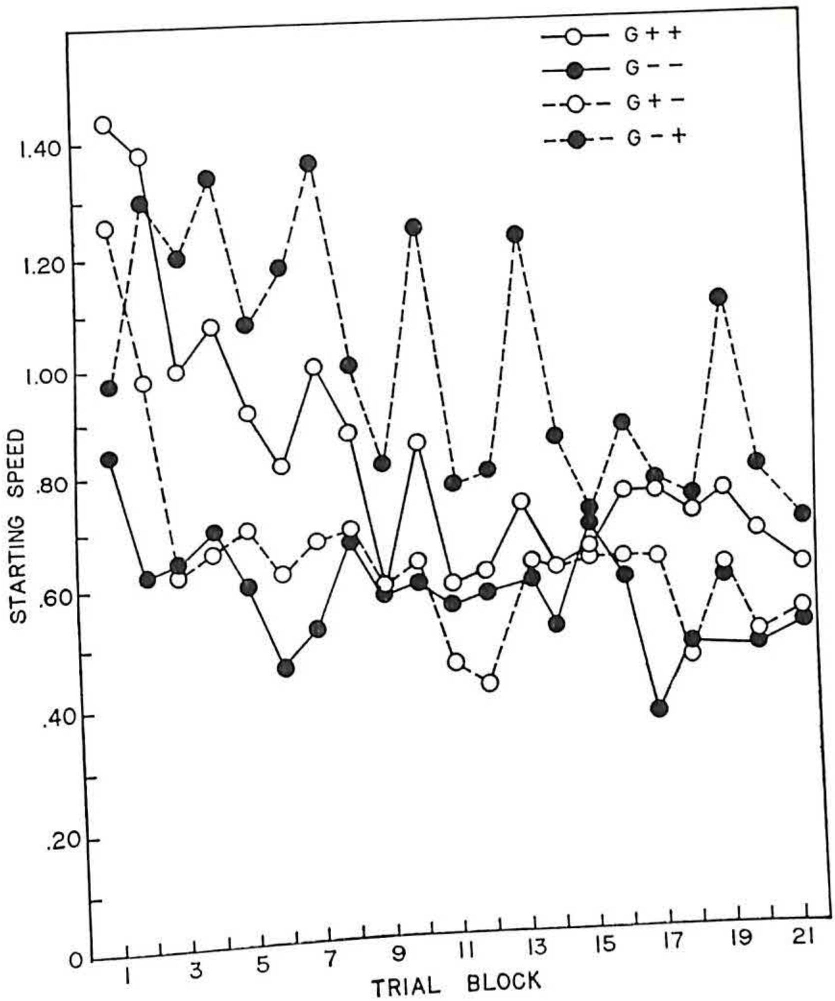


FIG. 3. Mean starting speed during extinction as a function of blocks of two trials.

× Trials interaction failed to prove reliable, F (51, 1326) = 1.17. When rate of extinction scores were computed, the group means were: Group +−, .274; Group ++, .437; Group −−, .462; Group −+, .964. Although these relationships were quite similar to those obtained for GB speeds, the overall differences in rate of extinction failed to prove reliable, F (3, 78) = 2.19, $.10 > p > .05$.

This failure to statistically confirm differences in rate of extinction among the groups appeared at least in part to result from the fact that the groups approached different asymptotes. Nevertheless, it may be cogently argued that, for example, Group ++ was more "resistant to extinction" than Group +−, since these groups entered extinction at same level of performance: for first trial, t (39) = 0.42, while Group ++ subsequently *maintained* its performance at a higher level, F (1, 39) = 5.20, $p < .01$. Similarly, Group −+ and Group −− entered performance at the same level, t (39) = 0.77, while Group −+ maintained its performance at a higher level, F (1, 39) = 9.73, $p < .001$. Finally, since Group −+ entered extinction at a lower level than Group ++, but subsequently performed at a higher level, F (1, 39) = 4.17, $p < .05$, it may be said to have shown greater resistance to extinction than Group ++.

Discussion

In general, the results for both starting and GB speeds indicate that Group −+ showed the greatest resistance to extinction and Group +− the least, while Group ++ and Group −− were intermediate. It is clear that certain aspects of these findings can be readily derived from the conventional concept of secondary reinforcement. Thus, when RW cues were equated (Group ++ vs. Group +−; and Group −+ vs. Group −−), the groups receiving secondary reinforcement in the GB showed greater resistance to extinction. On the other hand, when GB cues were equated (Group ++ vs. Group −+; and Group +− vs. Group −−), the secondary reinforcement effect failed to appear. In fact, those groups receiving secondary reinforcement in the RW clearly tended to show less resistance to extinction than the groups receiving a negative RW cue. Furthermore, these results cannot be interpreted as indicating simply that secondary reinforcement is effective only when presented in the GB, since the performance of groups receiving identical GB cues varied significantly as a function of RW cues. Rather, it was the *relationship* between RW and GB cues which constituted the most important determinant of extinction performance.

An alternative interpretation suggested by these data involves primarily the concept of incentive motivation. Spence (1956) has assumed that certain components (r_g's) of the consummatory $R(R_g)$ become conditioned to the GB on rewarded trials and, along with their afferent consequences (s_g's) generalize back into the RW and SB. The evocation of r_g-s_g is assumed to contribute to level of motivation in proportion to the vigor with which it is evoked. After training in a differential conditioning situation such as involved in the present study, r_g-s_g would presumably be vigorously evoked by the positive RW and GB, but only weakly evoked by the negative RW and GB. Further, Amsel (1958) and Spence (1960) have assumed that, once r_g-s_g has developed, nonreinforcement of the instrumental R results in the evocation of an emotional R, frustration, components of which (r_f-s_f) will also generalize to the RW and SB. During extinction following consistent reinforcement, frustration is assumed to elicit responses which compete with the instrumental R and hence, depress performance. The vigor with which such r_f-s_f is evoked is assumed to be a positive function of the strength of r_g-s_g.

The hypothesis proposed here is an

extension of this general line of reasoning. It is assumed that frustration occurs when an instrumental R is followed by a *decrement* in r_g-s_g and that the magnitude of the frustration response will depend on the rate and magnitude of reduction in r_g-s_g. It is further assumed that so-called "secondary reinforcement" occurs when an R is followed by an *increment* in r_g-s_g. "Primary reinforcement" in these terms would, of course, differ from "secondary reinforcement" only in the fact that the increment is in overt consummatory behavior (R_g), rather than r_g-s_g.

Application of these assumptions to the present results is facilitated by reference to Table 1, indicating the relative resistance to extinction in the comparisons of primary interest. Thus, the upper left-hand cell indicates that (*a*) Group ++ showed greater resistance to extinction than Group +−, (*b*) the RW cue was the same for these groups, and (*c*) that this result is presumably due to the greater frustration evoked following the instrumental R for Group +− than for Group ++. This last point follows from the present hypothesis in that r_g-s_g abruptly decreases for Group +− on opening the GB door, while for Group ++ the reduction occurs only after entering the GB, discovering the empty goal dish, etc. Furthermore, the magnitude of the decrement is presumably greater upon entering the negative GB than on entering the unbaited positive GB. The cells in the lower row of Table 1 indicate that Group −+ proved more resistant to extinction than either Group −− or Group ++. This finding is interpreted as a "secondary reinforcement" effect in the sense in which that concept is here employed—i.e., for Group −+ there is a greater increment in r_g-s_g upon entering the GB than for either Group ++ or Group −−.

TABLE 1

Expected Relationship among Groups

Source of Effect	Equated Cue	
	RW	GB
Frustration	G++ >G+−	G−− >G+−
Reinforcement	G−− <G−+	G++ <G−+

Note.—Cells indicate the group expected to show greater resistance to extinction. Columns indicate the cue on which incentive is equated, while rows indicate whether the expected difference reflects a frustration or a reinforcement effect.

It should be recalled that each expectation reflected in Table 1 was confirmed statistically for both starting and GB speed with the exception that the difference in *rate* of extinction for Group −− and Group +−, while in the expected direction for both starting and GB speeds, was significant only for the latter. It may also be noted that Group ++ and Group −− failed to differ in rate of extinction for both starting and GB measures. This finding is compatible with the present hypothesis, since, while r_g-s_g, presumably increases somewhat more for Group ++ than Group −− on entering the GB, the subsequent frustration for that group is also greater. These simultaneous and opposed processes might be expected to result in Group ++'s showing a rate of extinction comparable to that of Group −−.

Empirical support for the assumptions made above is not lacking elsewhere in the available literature. Thus, the findings that frustration appears increased following acquisition with large rewards, (Hulse, 1958; Wagner, 1961), following a large number of acquisition trials (North & Stimmel, 1960), and when nonreinforcement involves blocking completion of the instrumental R rather than simple omission of reward (Hulse & Stanley, 1956; Williams & Williams, 1943), are in each case implied by the hypothesis. Similarly, the facts that acquisition is facilitated by large reward and depressed by delay of reward follow from the assumption that magnitude of reinforcement (whether "primary" or "secondary") depends upon the rate and magnitude of increment in consummatory behavior following the instrumental R. The most interesting possibility, however, in the writer's opinion, is that the phenomena conventionally and circularly "explained" as "secondary rein-

forcement effects" may prove capable of systematic reinterpretation in terms of implicit consummatory behavior. Thus, the presentation of a stimulus previously paired with primary reinforcement (i.e., *overt* consummatory behavior) would be expected to strengthen an R—or forestall its extinction—because of its conditioned capacity to evoke *implicit* consummatory behavior, thus both lessening frustration and "secondarily reinforcing" the response in question.

REFERENCES

AMSEL, A. The role of frustrative nonreward in non-continuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

ANDERSON, N. H. Comparison of different populations: Resistance to extinction and transfer. *Psychol. Rev.*, 1963, **70**, 162–179.

BITTERMAN, M. E., FEDDERSEN, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

ELAM, C. B., TYLER, D. W., & BITTERMAN, M. E. A further study of secondary reinforcement and the discrimination hypothesis. *J. comp. physiol. Psychol.*, 1954, **47**, 381–384.

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

HULSE, S. H., & STANLEY, W. C. Extinction by omission of food as related to partial and secondary reinforcement. *J. exp. Psychol.*, 1956, **52**, 221–227.

NORTH, A. J., & STIMMEL, D. I. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227—234.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WAGNER, A. F. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WILLIAMS, S. B., & WILLIAMS, E. W. Barrier-frustration and extinction in instrumental conditioning. *Amer. J. Psychol.*, 1943, **56**, 247–261.

(Received December 12, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 75–83

A COMPARISON OF TWO PAYOFF FUNCTIONS ON MULTIPLE-CHOICE DECISION BEHAVIOR[1]

DAVID M. MESSICK AND AMNON RAPOPORT

University of North Carolina

An experiment was conducted to investigate the effect of 2 different payoff functions on the choice behavior of human Ss. The Ss, 36 undergraduates, were required to predict which of 10 stimuli would occur on each of 400 trials. One half of the Ss were paid according to an all-or-none payoff function, while the other half were paid according to a linear function. It was found that the relative frequency of response for the 10 alternatives differed markedly for the 2 groups. Moreover, it was found that Ss' behavior was compatible with the expected-utility hypothesis as opposed to the probability-matching hypothesis and a generalization of it. In addition, large and stable individual differences were found in terms of the expected value of Ss' behavior, V, and in terms of the entropy, H, of Ss' response distributions.

Interest in decision behavior in the two-choice situation was generated through the application of statistical learning theory to such tasks and through the discovery of the reliable probability-matching phenomenon. The behavioral law of probability matching is well established in the two-choice situation (Estes, 1957). In the three-choice situation with no monetary reward Komorita (1958) found the effect, but Gardner (1957) and Cotton and Rechtschaffen (1958) found reliable overguessing of the most frequent alternative. In addition, Gardner (1958) showed that the effect did not occur when the number of choices was increased. The relative frequency of response with the most frequent alternative tended to exceed the probability of that alternative, the difference increasing as the number of alternatives increased. Siegel and Goldstein (1959), Brackbill, Kappy, and Starr (1962), and Siegel and Andrews (1962) found that increasing the reward for making a correct choice and decreasing the reward for an incorrect one resulted in relative frequencies of responses with the most frequent alternative which exceeded the probability of that stimulus. Edwards (1956) moreover demonstrated that with an asymmetric payoff, the less likely of two events will be predicted more often than the more likely, if the payoff for a correct choice of the former is sufficiently larger than that for a correct choice of the latter.

Although statistical learning theory has been extended to account for the effects of magnitude of reward (Atkinson, 1962), some problems remain in the application of the theory to more complex decision contexts (e.g., Estes, 1962; Myers & Atkinson, 1963). An alternative theoretical approach exists which is general enough to cover the more complex situations. This is the decision-theoretic approach which postulates that human *S*s in multiple-choice situations execute strategies

[1] The authors are indebted to Lyle V. Jones for his helpful suggestions and critical reading of the manuscript. In addition the authors gratefully acknowledge the support of the Committee for the Support of Research and Advance Study, Executive Committee, General Federation of Labor, Tel Aviv, Israel, and of the United States Air Force Office of Scientific Research (Grant No. AF-AFOSR-62-5, and AF-AFOSR-85-63).

which tend to increase or to maximize *S*s' expected utility. The weakest statement of the above hypothesis (which will be called the Expected Utility—EU—hypothesis) is that the relative frequency of response with a particular alternative will be a nondecreasing function of the expected utility of the response. The expected utility maximization hypothesis, the relative expected loss minimization (RELM) principal (Edwards, 1956), and the models developed by Siegel (1959) are examples of EU models.

All of the experiments mentioned in the previous paragraphs are in accord with the general EU hypothesis. However, with the exception of the experiment by Edwards (1956), these experiments are also in agreement with a generalization of the probability-matching hypothesis which asserts that the relative frequency of response with a particular alternative is a nondecreasing function of the probability that the response will be correct (e.g., Atkinson, 1962). This is a result of the fact that these studies have been restricted to one specific type of payoff function, namely, an all-or-none payoff function in which a response on a given trial is either correct or incorrect. The present study is an attempt (*a*) to replicate the finding that the specific probability-matching phenomenon does not occur when the number of alternatives is large and when monetary payoffs are given for correct choices, and (*b*) to show that the EU hypothesis can be generalized to a different type of payoff function (i.e., a linear payoff function in which the payoff for a particular response is a linear function of the difference between that response and the response which would have been correct) while the generalized probability-matching hypothesis cannot.

Method

The *S*s were 36 undergraduate students enrolled in a summer course in introductory psychology. The *S*s were given course credit for their participation in the experiment. The *S*s reported individually to the dimly illuminated experimental room. They were divided into two groups as they appeared for the experiment. All odd-numbered *S*s, *S*-1, *S*-3, etc., were in Group 1 while even-numbered *S*s were in Group 2. The *S*s were seated 10 ft. from a 42 × 42 in. projection screen and read the instructions for the experiment.

The instructions which *S* read told him that his task in the experiment required him on each of 400 trials to guess how many 1's would appear in each of a set of 3 × 3 matrices composed of 0's and 1's which would be projected onto the screen in front of him. Only the instructions concerning the payoff differed for the two groups. In addition to the general instructions, *S*s in Group 1 were given the following "all-or-none" instructions.

> Your score for the experiment will be determined by the number of times you guess the slide correctly. Each time this occurs, you will receive a nickel. Try to win as many nickels as possible. All the money you win is yours.

The *S*s in Group 2 were given the following "linear payoff" instructions.

> Your score for the experiment will be determined by the difference between your guess and the actual number of 1's in the slide. You will receive payments for your guesses, depending upon your score. For example, if you guess "7" and the actual number is "4" the difference then is three. This difference will be subtracted from ten, and the remainder will be the number of pennies you win. In this case, you would receive 7 cents.

In order to make the average expected winnings of the two groups approximately equal, however, *S*s in Group 2 were told that they would be allowed to keep one tenth of their total winnings.

Formally, if X is *S*'s response and Y is the actual observed number of 1's in the 3 × 3 matrix, the all-or-none payoff function is represented by

$$P_a - \begin{cases} 5 \text{ cents, if } X = Y, \\ 0 \text{ cents, if } X \neq Y. \end{cases} \qquad [1]$$

The linear payoff is given by

$$P_b = \frac{1}{10}\left(10 - (|X - Y|)\right) = 1 - \frac{1}{10}(|X - Y|). \quad [2]$$

The strategy which maximizes the expected payoff in the all-or-none case is that of guessing the alternative with the highest probability of occurrence, in this case 0. The best strategy with the linear payoff is, in general, to guess the median value of the probability distribution. That value in this experiment is 3.

After having answered any questions, E suggested that S help him mix the slides to be used in the experiment and to place them in the carriage for the Kodak Cavalcade projector (Model 510). In all, 80 slides were used. The second and third columns of Table 1 give the number and percentage of slides, respectively, having n 1's in the 3×3 matrix, $n = 0, 1 \cdots 9$. The reason for using a stimulus distribution of this form is that such an inverse J-shaped distribution permits good separation of the modal and median values of the distribution.

The S was told to give his first prediction. Then the first slide was shown. This was repeated for 80 trials, with S giving his prediction on each trial and observing immediately the actual number of 1's in the matrix. In Group 1, each time S predicted correctly, he was told "You win a nickel," and a nickel was placed in a box near S. In Group 2, on each trial, E announced the difference between the prediction and the actual number of 1's and then announced the payoff. With both groups the total earnings were accumulated on a small adding machine. In neither group was S allowed to count his total earnings. The S was always given as much time as he desired to make his response. Average response time was about 5 sec.

After each block of 80 trials, it was suggested to S that he might help E reshuffle the slides and place them back in the carriage of the projector. This was done to emphasize further the random nature of the sequence and to eliminate the possibility that S might remember short sequences from the preceding block and use this information in making his guesses.

After 20 trials, S was asked to estimate the percentage of slides containing n 1's, $n = 0, 1 \cdots 9$. He was given a large (60×40 in.) board with 10 vertical grooves carved in it and a container holding 100 marbles. The vertical grooves were labeled 0, 1, 2, etc., from left to right. The S was reminded that each marble represented 1% and that he was to distribute the 100 marbles in the grooves in such a way that the number of marbles in each groove represented what he felt to be the percentage of slides with each specific number of 1's. Thus, if he felt that 25% of the slides had no 1's, he should put 25 marbles in the vertical groove labeled 0. The S was given as long as needed to do this task. After S was satisfied with his distribution, E recorded the number of marbles in each groove and scattered the marbles on the board in order to prevent S, the next time his estimates were called for, from seeing what he had done the previous time. These estimates of the proportions of the different stimuli were obtained after 20, 50, 80, 120, 160, 200, 240, and 400 trials.

TABLE 1

MEAN ESTIMATE OF PROPORTIONS OF STIMULUS ELEMENTS (NUMBER OF 1'S) AFTER 80 TRIALS AND AFTER 400 TRIALS

No. of 1's	No. of Slides with n 1's	True Proportion	After 80 Trials		After 400 Trials	
			Group 1	Group 2	Group 1	Group 2
0	18	.2250	.2033	.1717	.2306	.1750
1	11	.1375	.1239	.1006	.1167	.1056
2	8	.1000	.0856	.0933	.0872	.0994
3	6	.0750	.0883	.0872	.0744	.0928
4	5	.0625	.0678	.0767	.0883	.0733
5	4	.0500	.0656	.0800	.0617	.0694
6	4	.0500	.0583	.0733	.0578	.0678
7	5	.0625	.0694	.0839	.0633	.0839
8	7	.0875	.0706	.0789	.0789	.0928
9	12	.1500	.1672	.1544	.1711	.1400

TABLE 2

MEAN RELATIVE FREQUENCY OF RESPONSE FOR THE TWO GROUPS FOR EACH BLOCK OF 80 TRIALS

Blocks	No. of 1's Guessed									
	0	1	2	3	4	5	6	7	8	9
Group 1										
1	.1424	.0847	.1292	.1208	.0993	.0924	.0792	.0701	.0555	.1264
2	.2430	.1098	.0889	.0798	.0597	.0556	.0472	.0528	.0430	.2202
3	.3701	.1076	.0701	.0639	.0431	.0452	.0424	.0549	.0487	.1540
4	.3924	.0903	.0819	.0590	.0444	.0479	.0320	.0444	.0528	.1549
5	.4243	.0819	.0681	.0521	.0444	.0396	.0354	.0417	.0597	.1528
Group 2										
1	.0590	.0639	.1132	.1486	.1479	.1736	.1035	.0792	.0639	.0472
2	.0819	.0840	.1028	.1236	.1361	.1576	.0819	.0799	.0819	.0702
3	.1028	.0799	.1118	.1292	.1298	.1660	.0854	.0666	.0556	.0729
4	.1042	.0896	.1243	.1354	.1458	.1472	.0764	.0597	.0549	.0625
5	.0938	.0729	.1444	.1389	.1708	.1396	.0868	.0611	.0417	.0500

Finally, after Trial 400, *S* was asked, "If we were to continue and if you were allowed only one number as your prediction for the remainder of the experiment, which number would you select?" After responding, *S*s were questioned briefly to determine, if possible, what kind of cues or what type of strategy they employed in performing the task.

RESULTS AND DISCUSSION

The results of this experiment will be described in three sections. The first will deal with the relative accuracy of the generalized probability-matching model as opposed to the EU hypothesis in accounting for the observed data. The second section will be concerned more specifically with the evaluation of the EU hypothesis. The final section is devoted to a discussion of some highly stable individual differences found in the data.

Probability-matching vs. expected value.—The mean relative frequency of response for the two groups of 18 *S*s for each of the five blocks of 80 trials is given in Table 2. It is clear that in neither group is there evidence in the final block of trials of the specific probability-matching phenomena. It is further obvious that the responses of Group 2 (the linear payoff group) are not in accord with the generalized probability-matching hypothesis.

A statistical analysis was performed to verify this assertion. The statistical test employed is described by Page (1963). This test requires that the model specify a predicted ranking of the $k(k = 10)$ treatment means. Both the probability-matching model and the EU hypothesis predict such a ranking. With Group 1, however, both make the same predictions of the ranking of the relative frequencies of response, thus preventing discrimination between them in terms of ranks. The predicted rankings for Group 2 are quite different. The probability-matching model predicts, as in Group 1, that the relative frequency of response will increase as the probability of the corresponding stimulus increases. The EU hypothesis, on the other hand, predicts that relative frequency of response will be perfectly and positively correlated with the expected value of the response. Thus, for Group 2, a ranking is predicted by the EU hypothesis which is quite different from that predicted by the

generalized probability-matching hypothesis. Page's (1963) L statistic was computed for the data of Group 2 for each of the predicted rankings. A significant L indicates significant deviation of the data from the null hypothesis of no treatment differences in the direction of the ranking predicted by the model. The L computed for the ranking predicted by the probability-matching hypothesis is 5333.0, a value which is not significant at the .05 level. The L computed for the EU hypothesis ranking is 6000.5, which is significant at the .001 level. The data are consistent with the EU hypothesis and provide clear disconfirmation of the probability-matching hypothesis for Group 2.

It might be asserted that since the particular probability-matching model of Komorita states that S matches his *subjective* probability with the relative frequency of response and since the two payoff functions may systematically and differentially alter S's perception of the probabilities, then we do not have a direct test of this probability-matching model. It was partially for this reason that Ss were asked periodically to make estimates of the relative proportions of stimulus events which they had seen. The four rightmost columns of Table 1 show the mean estimate of the proportions of stimulus elements after the first 80 trials and after the final block of 80 trials for both groups. The actual proportion is shown for comparison in the third column.

The differences in the mean estimates after 80 trials and after 400 trials are too small to account for such a large difference between the response distributions of the two groups. Again, we find that the predicted relationship between subjective probability and relative frequency of response does not receive confirmation when the outcomes to Ss are determined by a linear payoff function.

The EU hypothesis.—As mentioned in the previous section, the L test indicated significant ($L = 6000.5$, $p < .001$) deviation from equality of relative frequency of response for different alternatives in the direction of the ranking predicted by the EU hypothesis. The L test was applied to the data of Group 1 using the ranking of relative frequencies predicted for this group by the expected-utility position. Again the results indicate significant deviation from the null hypothesis ($L = 6198.5$, $p < .001$) in the direction predicted. Thus the data from both groups are in accord with the EU hypothesis.

Although it is clear that the distributions of responses in the final block of 80 trials in Table 2 are quite different for the two groups, a rather weak statistical test was performed as verification and to provide an additional test of the EU hypothesis. The hypothesis states that Ss will tend to choose the response for which the expected value is greatest. This response is "0" in Group 1 and "3" in Group 2. For each S, the number of "0" responses, $f(0)$, and the number of "3" responses, $f(3)$, were counted in the last block of 80 trials. The difference, $d = f(0) - f(3)$ was found for each S in both groups, and the median of the distribution of d was found to be 9.5. The Ss were then classified according to whether the value of d was above or below the median. Of the 18 Ss in Group 1, 16 were above the median on d while 16 of the 18 Ss in Group 2 were below the median. This distribution of frequencies yields $\chi^2 = 16.9$, which, with $df = 1$, is significant at the .001 level.

The last source of evidence which is directly concerned with the EU hypothesis comes from S's responses

when asked which responses he would use if he could only use one. Sixteen of the 18 *S*s in Group 1 selected "0." The other two selected "9." (Both of these *S*s believed that there were as many "9" slides as "0" slides, as indicated by their estimates after 400 trials.) Only one *S* in Group 2 selected "0." Thirteen *S*s in Group 2 chose either "2," "3," or "4," 3 *S*s chose "5," and 1 chose "6." (Responses 2, 3, and 4 are combined since the difference in expected gain between guessing "3" every time for 80 trials and guessing "2"—or "4"—is less than 1 cent.) The difference between the two groups in the choices of a single response is unambiguous and provides clear support for the EU hypothesis.

Individual differences.—One of the most notable results of this experiment is that there are very large individual differences between *S*s in performing this task. For example, in Group 1, in the final block of 80 trials, the range of relative frequencies for responding with the most likely alternative, "0," is from .125 to 1.000. The individual differences were much more marked in *S*s' responses than in the obtained estimates, a result which certainly would be expected. With the estimates, *S* merely has to state his perception of the relative frequencies of the various environmental events. In making guesses, *S* must attempt to execute a strategy for maximizing the money he will win.

In order to investigate the consistency of these observed differences, a value measure, *V*, was computed for each *S* for each block of 80 trials. The formula for computing *V* is

$$V = \sum_{i=0}^{9} f(i)E_i(g), \qquad [3]$$

where i is the response variable going from 0 through 9; $f(i)$ is the frequency of response i in a block of 80 trials; and $E_i(g)$ is the expected gain for one response of i. This value is found for Group 1, by

$$E_i(g) = \pi(i)C, \qquad [4]$$

where $\pi(i)$ is the probability of stimulus event i, and C is the payoff for a correct response, in this case 5 cents. For Group 2,

$$E_i(g) = \sum_{j=0}^{9} \pi(j)(d - b|j - 1|), \qquad [5]$$

where summation is over all stimulus alternatives, $j = (0, 1 \cdots 9)$, d is the payoff for a correct guess (where $j = i$), and b is a scaling constant. In this experiment, $d = 1$ cent and $b = 10^{-1}$, thus a correct choice is worth 1 cent, a choice which is off by 1, $(|j - i| = 1)$ is worth .9 cents, etc.

After having computed *V* for each *S* for each block of 80 trials, the correlations of these values between all blocks of trials were computed, separately for each group. Table 3 shows the correlation matrices for Groups 1 and 2 along with the means and standard deviations for each block. Each correlation, mean, and standard deviation is based on data from 18 *S*s.

It is apparent that the differences between *S*s are very consistent. One fact which is of interest is that the correlations below the diagonal increase as the number of blocks increase. This is especially apparent in Group 1. In Group 1 the correlation of the value measure between Blocks 1 and 2 is .77, between Blocks 2 and 3 it is .89, between 3 and 4, .96, etc. This suggests that the individual response strategies are becoming more stable as the number of trials increases. (The large difference between the groups with respect to SD_V is a result of the fact that differences among $E_i(g)$ for the alternatives having greater expected values are much

TABLE 3

CORRELATIONS OF V BETWEEN BLOCKS, WITH MEANS AND SDs FOR EACH BLOCK FOR GROUPS 1 AND 2 SEPARATELY

Block	Correlations 1	2	3	4	5	$\bar{X}_V$	SD_V
Group 1							
1	1.00					43.50	9.18
2	.77**	1.00				53.44	11.72
3	.63**	.89**	1.00			59.44	16.28
4	.56**	.87**	.96**	1.00		60.29	15.44
5	.56**	.84**	.94**	.98**	1.00	61.89	15.87
Group 2							
1	1.00					52.56	1.17
2	.84**	1.00				51.87	1.79
3	.64**	.81**	1.00			52.06	1.92
4	.62**	.71**	.91**	1.00		52.37	1.87
5	.58**	.50*	.75**	.85**	1.00	52.77	1.58

* $p < .05$.
** $p < .01$.

greater under the all-or-none payoff function than under the linear payoff function.)

To investigate further the consistency of individual differences the uncertainty H of each S's response distribution for each block of 80 trials was computed by

$$H = -\Sigma P(i) \log P(i), \qquad [6]$$

where $P(i)$ is the relative frequency of responding with response i. This measure is related to V but not in a simple way. In general, maximum V and minimum V depend on H.[2] (This dependence will be noted by subscripting V, V_H.) H may be interpreted as a measure of the extent to which S distributes his 80 responses in a block over the set of all possible responses. The smaller the value of H, the more predictable S's responses. In our situation H can be as great as 3.322 if S guesses each alternative equally often, and as small as 0, if S guesses the same alternative for all 80 trials.

Individual differences between Ss with respect to H are highly consistent throughout the experiment. This is demonstrated in Table 4 which gives the correlations between H measures for all pairs of blocks of 80 trials, along with the mean and standard deviation of H for each block. The correlations for H are of the same order of magnitude as those in Table 3 for V.

In order to investigate the empirical relationship between V and H, the correlation between them in each block of trials was computed across Ss in each group. These correlations are given in Table 5. The fact that all of these correlations are negative indicates that increments in the value of Ss' behavior are accompanied by

[2] The mathematical problem of finding max V_H and min V_H is that of finding the maximum (minimum) of the equation

$$V_K = \sum_{i=0}^{9} a_i x_i,$$

where $a_i = E_i(g)$ and x_i is the relative frequency of response with the ith alternative, subject to the three constraints:

1. $x_i \geq 0$
2. $\sum_{i=0}^{9} x_i = 1.00$
3. $-\sum_{i=0}^{9} x_i \log_2 x_i = K$

TABLE 4

CORRELATIONS OF H BETWEEN BLOCKS, WITH MEANS AND SDs FOR EACH BLOCK FOR GROUPS 1 AND 2 SEPARATELY

Block	Correlations					$\bar{X}_H$	SD_H
	1	2	3	4	5		
Group 1							
1	1.00					2.964	.385
2	.62**	1.00				2.468	.671
3	.44*	.86**	1.00			2.325	.948
4	.36	.87**	.98**	1.00		2.308	1.006
5	.37	.84**	.96**	.98**	1.00	2.230	1.063
Group 2							
1	1.00					2.846	.489
2	.93**	1.00				2.569	.720
3	.94**	.97**	1.00			2.572	.797
4	.92**	.95**	.97**	1.00		2.533	.802
5	.83**	.91**	.93**	.95**	1.00	2.506	.922

* $p < .05$.
** $p < .01$.

decrements in the uncertainty of the behavior. These data are interesting from the point of view of the EU hypothesis, for if the relative frequency of response with an alternative is a nondecreasing function of the expected value of the alternative, then V_H will be very close to max V_H. It follows, in this case, that the only way to increase V is to decrease H. Inspection of the data reveals that this is, in fact, the process which is occurring. For most Ss the value V is close to or equal to max V_H. Thus the variation in H is interpreted as being responsible for much of the variation in V.

TABLE 5

CORRELATIONS BETWEEN TOTAL EXPECTED GAIN (V) AND RESPONSE UNCERTAINTY (H) FOR EACH BLOCK OF TRIALS

Group	Block				
	1	2	3	4	5
1	−.47	−.87	−.92	−.97	−.96
2	−.38	−.49	−.58	−.57	−.69

This interpretation of Ss' behavior places emphasis on individual differences with respect to H. The quantitative study of choice behavior and decision making has been little concerned with individual differences and as a consequence, scanty information is available regarding such differences. This fact, taken with the fact that these results were unexpected makes interpretation of these stable differences among Ss with respect to H somewhat speculative.

There is one obvious interpretation, however. It may be that these individual differences reflect differences in Ss' attitudes toward risk and gambling. The greater the H, in general, the greater the risk. If S guesses one alternative, i, all the time ($H = 0$), he guarantees himself that he will get $E_i(g) \cdot n$ where $E_i(g)$ is the expected value of one guess of the ith alternative and n is the number of trials. Deviations from an H of 0 increase the risk associated with the behavior.

REFERENCES

Atkinson, R. C. Choice behavior and monetary payoff: Strong and weak conditioning. In J. H. Griswell et al. (Eds.), *Mathematical methods in small group processes*. Stanford: Stanford Univer. Press, 1962.

Brackbill, Y., Kappy, M. S., & Starr, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63**, 32–35.

Cotton, J. W., & Rechtschaffen, A. Replication report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

Edwards, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

Estes, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

Estes, W. K. Theoretical treatments of differential reward in multiple-choice learning and two-person interactions. In J. H. Griswell et al. (Eds.), *Mathematical methods in small group processes*. Stanford: Stanford Univer. Press, 1962.

Gardner, R. A. Probability learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

Gardner, R. A. Multiple-choice decision-behavior. *Amer. J. Psychol.*, 1958, **71**, 710–717.

Komorita, S. S. Probability learning under equivalent data collection methods. *J. exp. Psychol.*, 1958, **55**, 115–120.

Myers, J. L., & Atkinson, R. C. Choice behavior and reward structure. Technical Report No. 54, 1963, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Page, E. B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *J. Amer. Statist. Ass.*, 1963, **58**, 216–230.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

Siegel, S., & Andrews, J. M. Magnitude of reinforcement and choice behavior in children. *J. exp. Psychol.*, 1962, **63**, 337–341.

Siegel, S., & Goldstein, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

(Received December 12, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 84–88

REVERSAL AND NONREVERSAL SHIFT LEARNING IN NORMAL CHILDREN AND RETARDATES OF COMPARABLE MENTAL AGE[1]

BARBARA SANDERS,[2] LEONARD E. ROSS, AND LAIRD W. HEAL

University of Wisconsin

48 normal and 48 retarded Ss were compared on reversal and nonreversal shift learning. Following the learning of the original problem, Ss were given training on 1 of 3 shift problems: Reversal (R), positive and negative cues reversed with 2 new irrelevant cues from the old irrelevant dimension introduced; Nonreversal Old (NRO), with previously irrelevant cues now relevant and 2 new cues from the old relevant dimension introduced as irrelevant; and Nonreversal New (NRN) the same as NRO except that the relevant cues were from the old irrelevant dimension but new. Reversal was significantly easier than either nonreversal condition for normals, but there were no significant reversal-nonreversal differences for retardates. NRO and NRN did not differ significantly for either group although both found NRO slightly more difficult. Retardates were significantly superior to normals in nonreversal and tended to be inferior in reversal.

Several investigators (Buss, 1956; Harrow & Friedman, 1958; Isaacs & Duncan, 1962; Kendler & D'Amato, 1955) have demonstrated that adults learn a reversal shift faster than a nonreversal shift. For nursery school children (Kendler, Kendler, & Wells, 1960) and for rats (Kelleher, 1956) the opposite has been found—a nonreversal shift is learned more quickly than a reversal shift. In a study with kindergarten children Kendler and Kendler (1959) divided their *S*s into a fast learning group and a slow learning group on the basis of the learning of the original problem and found that fast learners, like adults, found a reversal shift easier while slow learners, like nursery school children and rats, found a nonreversal shift easier. It was suggested that the fast learners responded in a manner consistent with a mediational S-R theory, while the slow learners performed as predicted by a single-unit S-R theory. Kendler and Kendler proposed that reversal would be easier than nonreversal for a majority of children past the age of about 6.

While the Kendler and Kendler prediction was made with respect to normal children, results in apparent agreement with it were reported by House and Zeaman (1962) for retardates of MA 6–8; i.e., these *S*s found a reversal problem easier than a nonreversal problem. This finding is surprising, however, if the postulated role of mediation processes in reversal and nonreversal situations is considered. Kendler and Kendler (1962), for example, speak of mediation as involving the implicit or

[1] Support was provided by Grant MH 06333-02 from the National Institute of Mental Health. The second author was a Kennedy Foundation Fellow at Peabody College during preparation of the report. The third author was supported by an NIH Predoctoral Fellowship. The authors express their appreciation to the following Wisconsin institutions and schools for their cooperation in providing *S*s and facilities: Bethesda Luthern Home, Watertown; Southern Wisconsin Colony and Training School, Union Grove; St. Colletta School, Jefferson; and Washington Elementary School, Madison.

[2] Now at Yale University.

explicit labeling of relevant and irrelevant stimulus elements. Retardates might be expected to have a verbal deficit, and therefore a mediation deficit, when compared to normals of similar MA. Even in cases where *S*s are matched on a verbal task the verbal responses of retardates might not serve a mediation function as well as those of normals, since it has been suggested (e.g., Luria, 1957) that the retardate deficiency is one of verbal control over behavior rather than a language deficit per se. If normal children learn a reversal shift faster than a nonreversal shift due to the utilization of mediation processes, and if retarded children suffer some sort of mediation deficit, then an interaction of IQ group by type of shift would be expected. That is, the retardates' reversal-nonreversal performance should be more similar, or performance should be reversed with nonreversal easier, when compared to the reversal-nonreversal learning of normals. The House and Zeaman results are not consistent with this line of reasoning since their MA 6–8 retardates performed as Kendler and Kendler would expect normals of that MA to perform. However the House and Zeaman results are difficult to interpret in this respect due to the absence of a normal comparison group, and procedural differences from studies using normals. In addition *S*s in their study received 125 overtraining trials on the original problem, a factor which could have served to retard nonreversal shift (Bensberg, 1958; Mackintosh, 1962).

The present study compared normals and retardates of similar mental age on reversal and nonreversal shift problems.

METHOD

Subjects.—The *S*s were 48 normal third-grade children and 48 mentally retarded children. Normals and retardates were of similar MA as determined by the Peabody Picture Vocabulary Test. Normals were from two public schools, retardates from a special public school summer session and Southern Wisconsin Colony.

Apparatus.—The apparatus was a modified Wisconsin General Test Apparatus, 30½ in. high and 24 in. wide, with 14½-in. wings slanting away from the front of the apparatus at an angle of 120°. Stimuli were introduced on a white plastic tray which was mounted on wheels and could be presented to *S* through a 7-in. high door below a one-way mirror. The door was closed between trials. Two circular hollows, 5½ in. apart, served as reward cups. Stimuli were slid over the cups in grooves on the tray in such a way that *S* was unable to lift a stimulus from the tray, but had to slide it away from him to expose the reward cup. The reward was a "cocoa puff" (General Mills' cocoa flavored cereal).

Stimuli were ¾-in. high wooden blocks mounted on 4 × 4¾ in. rectangles of white plastic. The shapes were not matched for area or volume, but were constructed so that there were margins of approximately ¼ in. between the edges of the shape and the edge of the plastic base. Forms were an equilateral triangle, a circle, a square, and a T. The triangle was presented with the apex toward *S*. Colors were red, blue, green, and yellow. In the original problem the red and the blue circle and triangle were designated Set 1 stimuli, and the green and the yellow square and T Set 2 stimuli. Four different stimuli composed of two forms and two colors were used for each *S*. These four stimuli were presented in pairs which varied simultaneously in form and color. (For example, for Set 1 stimuli the red circle was paired with the blue triangle for half the trials and the red triangle with the blue circle for the other half.)

Design.—The design was essentially a 2 × 3 × 2 × 2 × 2 × 2 factorial involving two IQ groups (retardates and normals), three conditions (R, NRO, and NRN), two dimensions in the transfer problem (color and form), two cue sets within each dimension, positive-negative (each cue of each set was positive once and negative once), and two sexes. Each of the 96 cells was filled by one *S*. Sex and positive-negative were often confounded in an effort to keep condition subgroups balanced with regard to MA, institutionalization, trials to criterion in the original discrimination, and number of *S*s requiring special training. There were 10, 9, and 11 institutionalized retardates in Retarded Subgroups R, NRO, and NRN, respectively.

Following a preliminary examination of the data only groups, conditions, and dimensions were used in the analysis, with all other factors collapsed. Thus for analysis there were 12 cells with 8 *S*s per cell.

After original learning each *S* was assigned to one of the following shift problems: *Reversal* (R): The previously negative cue became positive and the previously positive cue became negative. Two new irrelevant cues from the old irrelevant dimension replaced the old irrelevant cues. *Nonreversal Old* (NRO): The previously irrelevant cues became relevant, one positive and one negative. Two new cues from the old relevant dimension were introduced as irrelevant. *Nonreversal New* (NRN): The new cues from the previously irrelevant dimension became relevant, one positive and one negative. Two new cues from the old relevant dimension were introduced as irrelevant.

Two features of these conditions are noteworthy: (*a*) All conditions had new irrelevant cues in the shift problem, thus giving each *S* a "cue" to the change in the stimulus-reward contingencies and (*b*) the old relevant cues were removed for both nonreversal shift conditions thus precluding perseveration to the old positive cues.

Procedure.—Before the start of training *S* was given four trials without the stimulus blocks. For these trials the reward was placed alternately in the right and left reward cups. The baited cup was covered slightly more on successive trials with a white plastic rectangle that was identical to the stimulus bases. The last trial was given with the reward cup completely covered. On each trial *S* was required to slide back the cover, if necessary, and take the reward. Immediately after these four trials *S* was given the following instructions:

> Now I'm going to show you four blocks, two at a time. If you push back the right block there will be a cocoa puff under it; if you push back the wrong block there will be nothing under it. You may choose only one block each time. I want you to try to get as many cocoa puffs as you can.

Discrimination training was then started.

The order of presentation of the pairs of stimuli as well as the position of the positive stimulus on each trial followed a Gellerman sequence. A noncorrection method was used, and *S* was required to learn the problem to a criterion of 10 consecutive correct responses. If *S* failed to meet his criterion within 80 trials he was given the special training described below. However, if *S*'s last 7 responses in the 80 trials were correct he was given 3 additional trials to meet criterion. If *S* then failed to meet criterion after 83 trials the special training described below was given on the following day. Immediately after *S* met criterion on the original problem, and with no further instructions, he was given training on either R, NRO, or NRN. This training continued for 80 trials or until *S* met a criterion of 15 consecutive correct responses.

In order to avoid artifacts that would be produced by a high percentage of nonlearners, *S*s who failed to learn the original problem within 80 trials were given special training on the following day. The *E* first showed the positive stimuli to *S* saying, "These blocks are always right. There will always be a cocoa puff under these blocks." Then the negative stimuli were presented with the words, "These blocks are always wrong. There will never be a cocoa puff under these blocks." Training then proceeded as usual. Four *S*s who did not meet the criterion within the subsequent 20 trials were dropped from the study.

Two *S*s not available for training on the second day (e.g., absent from school) were discarded.

Results

The mean MA was 124.4 mo. for normals and 115.5 mo. for retardates. Although the groups differed somewhat with regard to mental age they were quite similar with respect to mean OL errors (23.86 for retardates; 23.40 for normals). Mean IQ was 70.7 for retardates and 110.4 for normals. Matching within groups on the basis of MA, IQ, and OL errors was such that in no case did the R, NRO, or NRN means within either group differ from the appropriate overall mean by more than one standard error. Assignment of special training *S*s resulted in 6, 9, and 7 normals; and 7, 5, and 7 retardates in Groups R, NRO, and NRN, respectively.

The mean number of errors made in reversal and nonreversal is shown in Fig. 1. A preliminary examination of the data revealed no main effects or interactions due to sex of *S*, and this factor was dropped from further analyses. A three-way analysis of

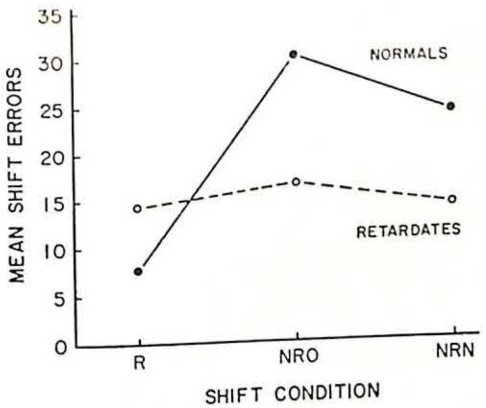


FIG. 1. Mean number of errors made by normals and retardates during reversal (R) and two nonreversal shift conditions (NRO and NRN).

variance of the error data involving groups (normal and retarded), conditions (R, NRO, and NRN), and dimensions (color and form) is summarized in Table 1. This table shows that, across groups, reversal *S*s made significantly fewer errors than nonreversal *S*s at the .01 level, with no significant difference between the two nonreversal conditions. The Groups × Conditions interaction was significant only in the comparison of reversal and nonreversal (R vs. NR). Figure 1 shows that the interaction arises from a marked decrement in the

TABLE 1

ANALYSIS OF VARIANCE OF ERRORS

Source	*df*	*MS*	*F*
Groups (G)	1	828.38	3.66
Conditions (C)	2	1419.60	
R vs. NR	1	2310.19	10.19**
NRO vs. NRN	1	529.00	2.33
Dimensions (D)	1	1837.50	8.11**
G × C	2	1044.03	
G × R vs. NR	1	2028.00	8.95**
G × NRO vs. NRN	1	60.06	.27
G × D	1	3.38	.02
D × C	2	340.72	
D × R vs. NR	1	609.19	2.69
D × NRO vs. NRN	1	72.25	.03
G × C × D	2	88.15	
G × D × R vs. NR	1	.73	.003
G × D × NRO vs. NRN	1	175.56	.77
Within	84	226.63	

** $p < .01$.

normals' performance on either nonreversal problem. Color was a significantly more difficult dimension than form ($p < .01$) but there were no significant interactions of the dimension factor with other variables.

Subsequent Duncan range tests revealed that the normal reversal group made significantly fewer errors than either normal nonreversal group ($p < .01$), but that there were no significant differences among the retardate groups. Duncan range tests were also used to compare normal and retardate performance under R, NRO, and NRN conditions. Only the NRO test reached the .05 level, although the NRN comparison approached that level.

DISCUSSION

The results of this study confirm the Kendler and Kendler (1959) prediction that a reversal problem is significantly easier than a nonreversal problem for a majority of normals past the age of 6. This prediction, however, cannot be extended to include retardates since the present study found no significant difference between reversal and nonreversal learning in retardates with a mean MA of approximately 9-8. These results are not in agreement with those of House and Zeaman, who found reversal learning significantly easier than nonreversal learning for retardates of MA 6–8, but are consistent with the notion that the mediation processes, in which the retarded are presumably deficient, are important in reversal-nonreversal situations. As previously mentioned, the use of an overlearning procedure may have been an important factor in the House and Zeaman results, although other procedural differences between the House and Zeaman study and the present experiment make consideration of the different results speculative.

In the present study there was no significant difference between NRO and NRN nonreversal problems for either normals or retardates, but both groups

found the NRN problem slightly easier. Familiarity with, or "learning to ignore" irrelevant cues may retard the subsequent learning of these cues to some degree, although for normals the reversal-nonreversal factor was of much more importance than the NRO-NRN procedure difference.

The present study indicates differences in the learning processes of normals and retardates of approximately the same mental age. While the normals' performance was such as to suggest that mediation processes were utilized (reversal easier than nonreversal), the retardates' performance did not indicate a similar involvement of mediation. However, if the retardates had been completely unable to utilize mediational processes, reversal should have been more difficult than nonreversal, as has been found in studies with young children and rats. These results might, of course, be due to the presence in the retardate group of some *S*s who utilized mediation processes and some who did not. Such an interpretation of similar reversal-nonreversal results was proposed by Kendler and Kendler (1959), who then related the mediating behavior of their *S*s to performance on the original problem. A comparison of the shift performance of fast and slow learners was not possible with the present paradigm since errors in original learning could not be separated from a dimension preference. If original learning errors reflected the chance assignment of *S*s to preferred or nonpreferred dimensions—the fast learners having been assigned to their preferred dimensions and the slow to their nonpreferred dimension—errors in original learning would be expected to correlate positively with reversal shift errors and negatively with nonreversal errors. Correlations between original learning errors and shift errors were in these directions. The *S*'s "degree of mediation" could not be related to MA, IQ, or CA since all of these measures failed to correlate significantly with number of errors on the nonreversal shift problem.

REFERENCES

Bensberg, G. J. Concept learning in mental defectives as a function of appropriate and inappropriate "attention sets." *J. educ. Psychol.*, 1958, **49**, 137–143.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–598.

House, B. J., & Zeaman, D. Reversal and nonreversal shifts in discrimination learning in retardates. *J. exp. Psychol.*, 1962, **63**, 444–451.

Isaacs, I. D., & Duncan, C. P. Reversal and nonreversal shifts within and between dimensions in concept formation. *J. exp. Psychol.*, 1962, **64**, 580–585.

Kelleher, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

Kendler, T. S., & Kendler, H. H. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

Luria, A. R. The role of language in the formation of temporary connections. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957.

Mackintosh, N. J. The effects of overtraining on a reversal and a nonreversal shift. *J. comp. physiol. Psychol.*, 1962, **55**, 555–559.

(Received December 16, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 89–91

INCREASING CREATIVITY BY FREE-ASSOCIATION TRAINING

JONATHAN L. FREEDMAN[1]

Stanford University

An experiment was conducted to test the hypothesis that facilitating S's ability to produce associations would increase his score on a test of creativity. Ss either free associated to 10 stimulus words or defined the words. Those receiving the free-association training scored significantly higher on the Remote Associates Test (RAT), a test of creativity which was administered after the training session.

Maltzman (1960) has proposed that creative thinking consists of bringing together various associations which are novel but are also appropriate to the task. To test this idea he and his associates have conducted a series of studies (Maltzman, Bogartz, & Breger, 1958; Maltzman, Simon, Raskin, & Licht, 1960) in which *S*s free associate to a series of stimulus words under a variety of conditions and then take the Unusual Uses Test (Guilford, 1950). The one procedure which has been consistently successful in improving performance on the Unusual Uses Test (as compared with control groups) is presenting the stimuli a number of times and requiring *S*s to give a different response to each stimulus on each presentation.

A recent experiment by Caron, Unger, and Parloff (1963) indicated, however, that this training procedure did not affect performance on the Remote Associates Test (RAT). Since Mednick (1962) has shown that *S*s who have a relatively large number of associations to a given stimulus tend to score higher on the RAT than those who have fewer associations, a training technique which facilitated free associating should improve performance on the RAT. The failure of Maltzman's procedure may have been due to the fact that it is designed to increase the production of unusual associations, whereas the RAT is constructed so that the correct responses are commonly associated with the stimulus words, and are often dominant associates (e.g., MOUSE-CHEESE). Since it appears that success on the RAT depends upon the production of many associations in a short time, the present study attempted to improve performance on the RAT by a training technique specifically designed to facilitate this process.

METHOD

Subjects.—The *S*s were 90 students from an introductory psychology class at Stanford University. There were 20 males and 20 females in each of the two main experimental conditions, and an additional 10 males in a control condition. The *S*s were run individually.

Procedure.—All *S*s were told that they were going to take a test designed to measure creativity and that before the test they would receive a brief warm-up exercise. The facilitation group was instructed to give their associations to each of 10 stimulus words. Each word was read aloud by *E* and *S* was told to give whatever words came into his mind, and to continue until told to stop. They were stopped after 30 sec. or when they had stopped responding. The nonfacilitation group was asked to define each of the words. The same words were used for both groups, and were chosen so that neither the words themselves nor any of their common associates were the correct responses to any of the items on the creative thinking test. The

[1] The author is grateful to Zeta Brown who assisted in the running of this study.

stimulus words were: BOOK, MOON, COPPER, MOVIE, BRASS, COFFEE, READ, FARM, TOGETHER, STAR.

An additional control group consisted of 10 male *S*s, each of whom was paired with 1 male *S* in the facilitation condition. The procedure was identical to that in the facilitation condition except that the control *S*s read from a card the associations which had been given by the paired experimental *S*. Thus, in each pair both *S*s had verbalized the same words, but the facilitation *S* had produced them himself, whereas the control *S* merely read them.

The creative thinking test was the Remote Associates Test (RAT). This test is composed of items consisting of three words, and *S*'s task is to find a fourth word which is somehow associated with each of the three stimulus words. The test consists of 30 items and takes 40 min. It was presented in two equal parts of 15 items, and the original warm-up exercise was repeated after the first part of the test. Scores on the test consist simply of number correct. (For a more complete description of the RAT see Mednick, 1962.)

A pretest similar to the present study indicated that the experimental manipulation was more successful with men than with women. Since the pretest was run entirely by a male *E*, it seemed likely that female *S*s were constrained in free associating in the presence of a male *E*. Therefore, in the present study all men were run by a male *E* and women by a female *E*.

Results and Discussion

The major results are presented in Table 1, which shows the score on the RAT for facilitation and nonfacilitation groups and for male and female *S*s. The means for the complete test are presented because the results are similar for the two halves of the RAT. For both men and women the facilitation group achieved higher scores than the nonfacilitation group. As may be seen in the accompanying analysis of variance (Table 2), the difference between facilitation and nonfacilitation groups is significant. Separate comparisons using the combined error term indicated that the difference is also significant for males alone; $F\ (1, 76) = 10.25$, $p < .01$, but not for female *S*s, $F\ (1, 76) = 1.52$.

TABLE 1

Mean Scores on RAT

	Males	Females
Facilitation	18.05[a]	18.60
Nonfacilitation	14.15	17.10

[a] Maximum score is 30.

TABLE 2

Analysis of Variance of Mean Scores on RAT

Source	*df*	*F*
Facilitation-Nonfacilitation (F-NF)	1	9.82**
Male-Female (M-F)	1	4.13*
F-NF × M-F	1	1.94
Error (*MS*)	76	(14.84)

* $p < .05$.
** $p < .01$.

It is difficult to determine why the effect is so much stronger for male than for female *S*s. There is, of course, a confounding of *E* and sex of *S*s, but an extensive pretest using a single *E* produced essentially the same pattern of results. There were no appreciable differences in the number of associations produced by the two groups during the training session. It is possible that the difference is due to a ceiling effect for female *S*s since the overall means for females are significantly higher than for males.

It could be argued that the facilitation *S*s scored higher on the RAT because during the training session the former produced a large number of associations to the stimulus word whereas the latter merely defined it. Although the associations were not the exact responses necessary for the RAT, they may have indirectly strengthened the correct responses through the process of mediated generalization (Mednick & Freedman, 1960). The control group was run to assess this possibility.

The control *S*s had available to them the same body of words that were available to the facilitation *S*s, but the former had not actually free associated. If the availability of the words were the critical factor, the control group should have scored as high as the facilitation group on the RAT; if the act of associating were critical, the control group should score less high and should not differ from the nonfacilitation group. The latter result was found. The control group mean was 14.56 which was significantly lower than the paired facilitation group as indicated by a test for correlated means; $t(9) = 2.31$, $p < .05$, and did not differ appreciably from the nonfacilitation group. Thus, it appears that the actual act of associating is the critical factor which facilitated performance on the RAT.

Why did this training procedure have a facilitating effect when Maltzman's technique did not? The explanation does not seem to be a difference in the amount of training since the present procedure seems to have given less training than the standard Maltzman technique. The number of responses per stimulus was fewer in the present study (a mode[2] of three vs. six in Caron's study); fewer stimuli were used (10 vs. 25); and the total time taken by the training was considerably less (about 5 min. vs. about 20 min.).

There was probably a difference in the remoteness of the associations given during training in the two studies. Unfortunately, it is impossible to assess the remoteness of the responses given in the present study since no appropriate norms exist. The *S*s did not give a number of responses to the same stimulus, but were giving a series of associations with, in most cases, only the first one being specifically to the stimulus word. It may be that this gives *S*s practice associating to several stimulus words at a time, which is what he must do on the RAT.

Another possibility is that, as discussed earlier, the present technique causes *S*s to produce a number of associations in a brief period whereas the Maltzman method requires only one association at a time. Since *S*s probably solve the RAT problems by trying to produce many associations quickly, the present training technique is more similar to the process of solution than is Maltzman's; and the former might therefore be expected to have a greater facilitating effect than the latter.

[2] The mode is probably the most descriptive measure of central tendency since a large majority of *S*s gave exactly three responses to each stimulus but a few gave a very large number of responses.

REFERENCES

Caron, A. J., Unger, S. M., & Parloff, M. B. A test of Maltzman's theory of originality training. *J. verbal Learn. verbal Behav.*, 1963, **1**, 436–442.

Guilford, J. P. Creativity. *Amer. Psychologist*, 1950, **5**, 444–454.

Maltzman, I. On the training of originality. *Psychol. Rev.*, 1960, **67**, 229–242.

Maltzman, I., Bogartz, W., & Breger, L. A procedure for increasing word association originality and its transfer effects. *J. exp. Psychol.*, 1958, **56**, 393–398.

Maltzman, I., Simon, S., Raskin, D., & Licht, L. Experimental studies in the training of originality. *Psychol. Monogr.*, 1960, **74**(6, Whole No. 493).

Mednick, S. A. The associative basis of the creative process. *Psychol. Rev.*, 1962, **69**, 220–232.

Mednick, S. A., & Freedman, J. L. Facilitation of concept formation through mediated generalization. *J. exp. Psychol.*, 1960, **60**, 278–283.

(Received December 19, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 92–99

PREDICTION ON THE BASIS OF CONDITIONAL PROBABILITIES [1]

WILLIAM C. HOWELL AND JOSEPH F. FUNARO

Ohio State University

2-choice prediction behavior was studied for a task in which outcome probabilities were conditional upon values of the stimuli. S was required to guess which of 2 meter readings would be higher given one as a stimulus; readings were drawn in pairs from 2 separate distributions. 4 groups of Ss made 100, 300, 500, or 700 predictions, following which they judged distribution means and various probabilities associated with the readings. Asymptotic prediction frequencies did not match conditional probabilities even though judgments were surprisingly accurate. Prediction functions suggest that S establishes cutoff points for each stimulus distribution beyond which he chooses the alternative meter. Empirical cutoff data support this interpretation suggesting, further, that the location of cutoff points approaches the distribution means as knowledge is gained over trials.

The manner in which *S* deals with the classical, two-choice, probability-learning situation is fairly well established at an empirical level (Edwards, 1961; Neimark & Shuford, 1959; Restle, 1961; Siegel & Goldstein, 1959). Although disagreement exists as to whether the underlying mechanisms involve probability matching, schema formation, extreme-asymptote generalization, expected-utility maximization, or some combination thereof (Edwards, 1961; Estes, 1950; Grant, Hake, & Hornseth, 1951; Restle, 1961; Siegel, 1959), it is apparent that *S* can—and under some conditions does—produce response frequencies which parallel closely the probabilities of alternative stimuli. Such probability matching has even been reported for situations in which the choice required "probabilistic discrimination" (Estes, Burke, Atkinson, & Frankmann, 1957), a finding which supports nicely the model proposed by Burke and Estes (1957). The basis for discrimination in this study was the frequency with which specific stimulus elements (lights) appeared in samples drawn from a set of 12 such elements. Therefore, any given stimulus sample was representative of one of two sets of sampling rules and each response was reinforced in accordance with the rule in effect at the time.

A slightly different problem involving probabilistic discrimination is one in which reinforcement is administered on the basis of *direct comparisons* of *individual* stimulus elements (herein termed *values*) representing the two sampling processes. The present study employs a task which permits the investigation of this problem. As in the case of the Estes et al. study, the basic difference between this and the classical task is that the alternative stimuli (A and B) assume any of

[1] This research was carried out in the Laboratory of Aviation Psychology and was sponsored by the Advance Research Projects Agency, Department of Defense, under Contract No. AF 33(657)-11718 with the Ohio State University Research Foundation, and monitored by the Aerospace Medical Division, Air Force Systems Command, United States Air Force. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part for any purpose of the United States Government.

a number of values rather than merely occurrence vs. nonoccurrence on any given trial. Here, however, S is provided with the value (reading) for only one stimulus and is required to guess whether it or the other stimulus (not given) has the higher reading. On successive trials, pairs of A and B readings are drawn randomly from separate A and B populations; thus, S can benefit from past experience by formulating increasingly reliable estimates of the distribution parameters.

The present task permitted the investigation of several related problems. The first objective was to determine the extent to which asymptotic frequencies of choice at each value of A and B match their conditional probabilities—i.e., the probabilities of A or B being greater given (*a*) a specific value of A, or (*b*) a specific value of B. As in the classical problem, such behavior would not tend to optimize correct guesses, although it would be precisely the behavior predicted by Estes' (1950) theory. Evidence for conditional probability matching is, in fact, provided by the Estes et al. (1957) study (see Fig. 3). A second objective was to determine the manner in which behavior changes over trials; since the present task involves a large number of conditional statements rather than a pair of unconditional ones, it would seem that asymptotic performance would be reached more slowly than in the classical situation. A final objective was to determine, insofar as possible, what it is that S acquires over trials. Neimark and Shuford (1959) broached this issue for the classical problem by attempting to relate estimates of probability to choice frequencies. Unfortunately, the operation of estimating appeared to influence that of predicting so that the results on this point were inconclusive. In the present study the approach was to use separate groups for the various (four) levels of practice with each group making estimates following the completion of their trials. As will be discussed below, estimations were obtained for a number of distribution characteristics.

METHOD

Subjects.—The Ss were 24 male and 16 female undergraduate students enrolled in an introductory psychology course. They were divided at random into four groups of 10 each (6 males and 4 females).

Apparatus.—The display consisted of a 10×24 in. panel upon which was mounted a pair of microammeters with scale markers reading from 0 to 200 μa. in units of 5. The left meter was labeled A and the right one B. The meters could be set independently to any desired reading by adjusting potentiometers in circuits from a power supply to each meter. A switch in each circuit afforded E manual control over the presentation of readings on one or both meters. Another pair of switches was located beneath the display panel for use by S in making his response. His choice of meters was indicated to E by an A or B light which was activated by the A or B response switch. In addition, *either* response switch made the circuits from the power supply to *both* meters so that S, in responding, would get feedback from both meters (i.e., he would see both the A and the B readings for that trial).

The S was seated at a table upon which were mounted the display and corresponding response keys. A partition separated him from E whose apparatus was located on the table behind the partition. On each trial E made settings on both potentiometers, gave a verbal ready signal to S, closed the switch to *one* of the meters (i.e., presenting a reading), and waited for S's response. The S then pressed either the A or B response switch according to which meter he believed was set to the higher reading. His choice was signaled to E by a light and the readings on *both* meters were displayed as feedback. Release of the switch by S terminated the trial. Approximately 4 sec. intervened between trials during which time responses were recorded and settings were made.

Procedure.—As noted above, the task required S to guess which of the two meters on his display had the higher reading given the

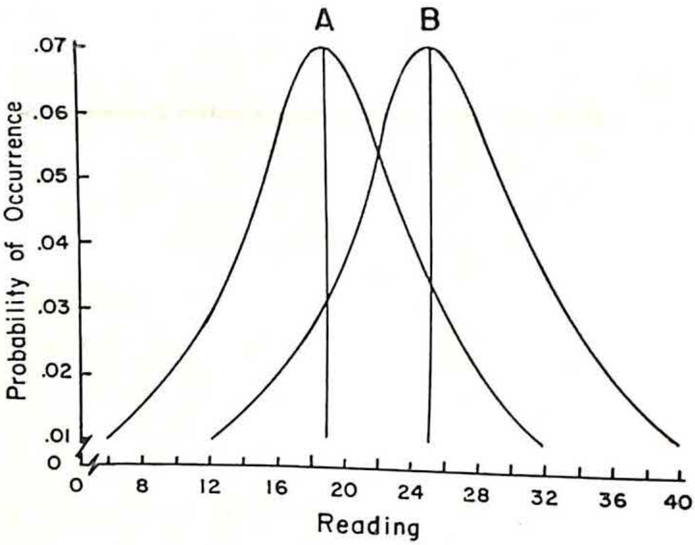


FIG. 1. An illustration of the distributions of readings sampled on Meter A and Meter B.

reading for only one of them (the stimulus reading). On successive trials the stimulus reading appeared alternately on Meter A and Meter B so that S would have equal opportunity to sample both meters. After each response, of course, S was presented both readings (A and B) as knowledge of results.

The readings to be used on any trial were drawn at random and independently from two normal populations: one from the A population and one from the B population. These populations of readings were defined, as shown in Fig. 1, with means of 19 and 25 and 's of 6 units (hence the means differed by 1 σ). nstructions contained no mention of the xistence or nature of the underlying populaions; a major purpose of the study was, in act, to determine how much S learned about hem and to what extent he reflected his nowledge in his behavior. One restriction laced upon the sampling procedure was that here be no tied readings; whenever this ocurred, that pair was discarded and a new air was drawn. All pairs for a given session ere entered on program sheets for E's use in anual presentation of the readings.

Trials were administered in blocks of 50 th summary feedback (number of errors) ovided after every 100 trials. Groups were fined by the number of trials performed: oup I—100; Group II—300; Group III—); Group IV—700. In each daily session e the first, 200 trials were completed; in first session only 100 were administered. llowing the completion of all trials, each S s required to estimate (a) the average readings for A and B, (b) the number of times in 100 that each of 12 readings from each distribution would occur, (c) the number of times in 100 that Meter A (or B) would read higher given each of 12 stimulus readings on each meter, (d) the number of times in 100 that Meter B would read higher than A. For Questions b and c above, the readings concerned were actually presented on the display. All instructions for this portion of the study were printed on sheets for S to read and were also presented verbally by E to insure complete understanding. Verbal responses were recorded by E.

RESULTS

The statistical nature of the task, as illustrated in Fig. 1, was such that S could optimize correct predictions by adopting the following rule: choose A only if A, when presented, reads higher than the mean of B or if B, when presented, reads lower than the mean of A; otherwise, choose B. The extent to which performance approaches this optimum can be estimated by calculating the empirical cutoff points at which S switches his prediction from the meter on which a reading is presented to that on which it is not. That is, an estimate of the A distribution cutoff is defined as the

point midway between the highest B value at which A is chosen and the lowest B value at which B is chosen; for the B distribution, it is midway between the highest A value at which B is chosen and the lowest A value at which A is chosen. Such cutoff points were determined for all groups on the last 100 trials, and the means are shown in Fig. 2. One set of curves in the figure reflects the performance of *different* groups at various levels of practice, whereas the others are based upon a single group (IV) at the same levels. The optimum cutoff points, i.e., distribution means, are also plotted for purposes of comparison. Over the 700 trials it is clear that the cutoff points diverge and approach an asymptote near—but not on—the optimum levels. A t test revealed that the difference between A and B cutoffs for Group IV at 700 trials is highly significant ($p < .01$). The σ_m for the B cutoff scores (Group IV—700 trials) was .47 and for the A cutoff scores was .62. Therefore, the B distribution mean lies within the .05 confidence interval of the B cutoff scores, whereas the A distribution

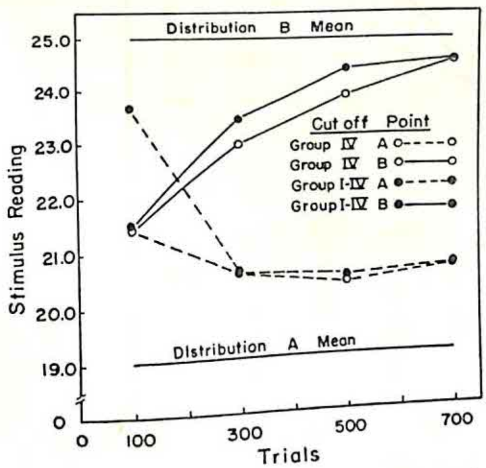


FIG. 2. Mean cutoff points obtained for the two distributions plotted for a single group and the four separate groups after comparable periods of practice.

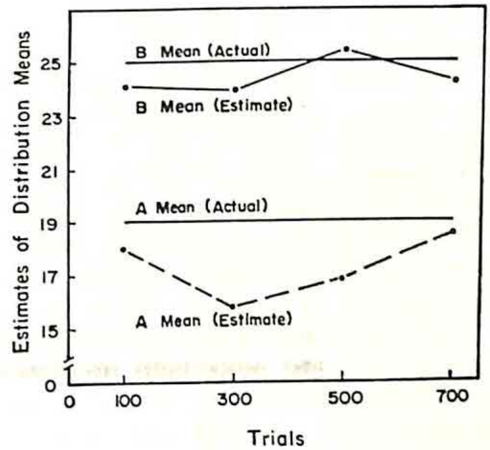


FIG. 3. A comparison of mean estimates with actual mean values for both distributions at all levels of practice.

mean lies outside the .05 confidence interval for the A scores. This means that B cutoffs cannot be regarded as significantly different from their optimum level whereas A cutoffs can.

Figure 3 summarizes the average estimates of distribution means for A and B readings made by *S*s at varying levels of practice. Here, of course, trials and groups are completely confounded. An analysis of variance carried out on these data indicated that the A vs. B effect is significant ($p < .01$) but the practice (group) effect is not ($p > .05$). This suggests that *S* develops a fairly good estimate of the two means in 100 samples (trials), which estimate improves little—if at all—over the next 100 trials.

In contrast to mean judgment, estimates which are intended to reflect *S*s' knowledge of the *form* of the A and B distributions were too unreliable to be of any use. The responses of some *S*s suggested that they judged, instead of the probability of *occurrence* of various readings, the conditional probability of a meter being *correct* given these readings. Such judgments were in marked contrast to those of other

*S*s who apparently interpreted the instructions correctly.

The extent to which *S* responds in accordance with conditional probabilities is described in Fig. 4. Here, the response proportions for the B meter are compared with the probabilities of the B meter reading higher when various A and B readings are presented. In order to obtain more reliable estimates of both probabilities and proportions (especially for low-probability events), the data are grouped over *S*s and over limited ranges of readings. Response proportions are based upon the last 100 trials for each group and conditional probabilities upon the A and B stimulus populations. To simplify the presentation, only the data for Group I and Group IV are included; those for the other two groups lie well within these limits and are very similar in form to Group IV. It is clear that none of the functions in Fig. 4 are described adequately by the diagonal line which represents the probability-matching hypothesis. The data for 100 trials (Group I) are obviously closer to this hypothesis than those for 700 trials (Group IV); with practice, apparently, the function becomes increasingly sigmoid. In neither group do there appear to be any systematic differences attributable to the stimulus. It should be noted, in regard to these functions, that the maximum number of correct guesses would be obtained if the proportion of B choices dropped directly from 1.00 above an actual probability of .50 to .00 below this point. Such a

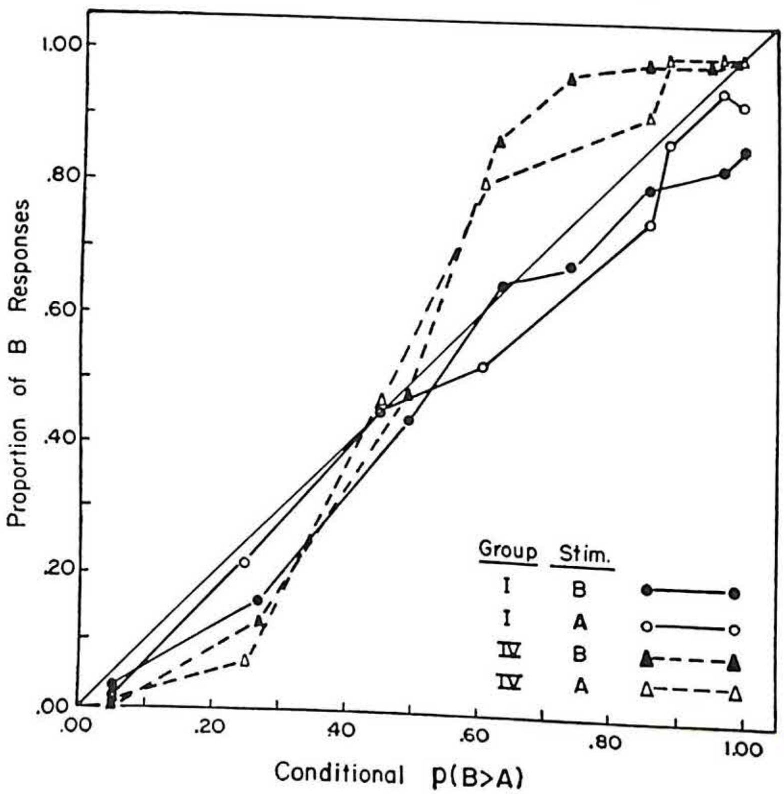


FIG. 4. Proportion of times Meter B was chosen over Meter A given A or B readings with different conditional probabilities of B > A; Group I (100 trials) and Group IV (700 trials) compared.

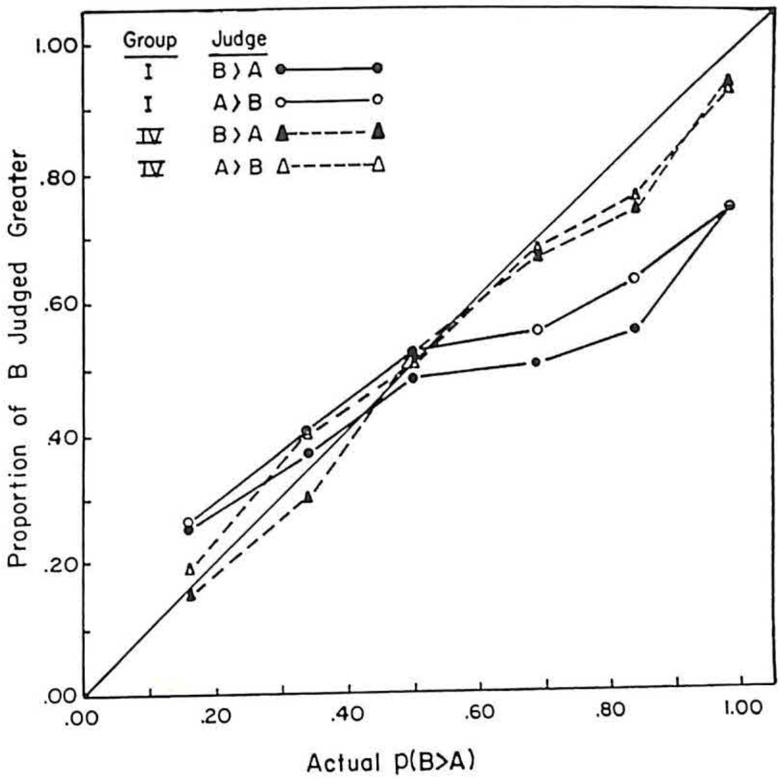


FIG. 5. Median proportion of times Meter B was judged greater than Meter A given readings with various conditional probabilities of B > A; Group I (100 trials) and Group IV (700 trials) compared.

function would result if S were to adopt the strategy described earlier in which cutoffs are located at the A and B distribution means.

Turning to the data which directly reflect knowledge gained over trials, Fig. 5 indicates that S does, indeed, come to achieve a fairly good understanding of the conditional probabilities involved for the various stimulus readings. Again, data are plotted only for the extreme levels of practice (100 and 700 trials), and it is clear that judgments are far closer to actual probabilities on Trial 700 than on Trial 100. On the earlier trials, S appears to underestimate high probabilities and overestimate low ones. Even after 700 trials there appears to be slight tendency to underestimate high conditional probabilities.

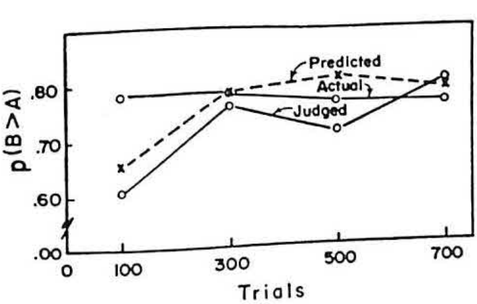


FIG. 6. A comparison of the median proportions of times Meter B was judged, was predicted, and was actually greater than Meter A at four levels of practice.

The last judgment required of S was that of the simple probability of the B meter reading higher than the A meter (irrespective of stimulus). As shown in Fig. 6, these judgments improved between 100 and 300 trials but little thereafter. An interesting par-

allel is shown in this figure between probability judgments and proportions of B choices on prediction trials. Both functions approach an asymptote after 300 trials at a level very close to the *actual* proportion of times B is greater than A.

Discussion

The present results indicate that asymptotic response frequencies do not match conditional probabilities associated with the stimuli in a complex version of the two-choice guessing game and, further, that the discrepancy becomes greater as *S* gains knowledge of the probabilities involved. It is equally clear that this behavior is not attributable to *S*'s failure to learn the statistical parameters of the task: the data reveal that he judges the mean, the simple $B > A$ probability, and even the conditional probabilities ($B > A|a$, $B > A|b$, $A > B|a$, $A > B|b$) with considerable accuracy after several hundred trials. He simply prefers, it seems, to make his choices on some basis other than the maladaptive probability-matching principle. The nature of his strategy, while not revealed unequivocally by the data, appears to be one intended to maximize the number of correct guesses: i.e., choose a stimulus reading for both meters below which you will predict the alternative meter to be greater; set these points at readings for which the conditional probability of success is .50 (in this case, the mean of the distribution of readings on the other meter). Such a strategy is essentially that proposed by statistical decision theorists to account for a wide range of behavioral phenomena (Swets, Tanner, & Birdsall, 1961). Figure 4 indicates that asymptotic choice of the B meter does, indeed, approach 1.00 as the conditional probability of $B > A$ increases above .50. That it does not actually become a step function, as would be the case if *S* were to maximize correct guesses, is partially due to the fact that data are combined over *S*s. So long as there are individual differences in the choice of cutoff points, such treatment must result in a degree of curvature; so long as there are individual differences in the judgment of means, it is probable that cutoff points will vary. From the data summarized in Fig. 3, it is apparent that there is still *some* constant and variable error in mean judgments even after 700 trials.

It should be noted that the present results are at variance both with the Burke-Estes (1957) theory and with the empirical data supporting it (Estes et al., 1957). In the latter study, the response probability to each of 12 stimulus elements was ". . . virtually equal to the conditional probability of reinforcement [p. 236]." The discrepancy between the Estes results and those of the present study suggests that reinforcement based upon direct comparison of elements from the two sampling processes (as in the present study) leads to the more rapid acquisition of an optimal choice strategy. Thus, the less-optimal, probability-matching strategy reported by Estes would be expected ultimately to be replaced by the more-optimal decision strategy observed in the present study if training were continued. Just such a transition has been reported by Edwards (1961) for the simple two-choice situation.

The frequency with which the B meter is chosen as greater than A, regardless of reading, approaches an asymptote at 300 trials (Fig. 6). This is also consistent with Edward's (1961) findings for the simple two-choice situation. It is quite apparent, however, that changes in behavior continue well beyond 300 trials as illustrated by the cutoff data in Fig. 2. Most of the improvement is with regard to the location of the B distribution cutoff point which, it will be recalled, is estimated from the magnitude of A values required to effect a choice of the B distribution. Since A values are infrequently larger than B values, and hence receive fewer reinforcements when presented, it is not surprising that *S* should require more trials to reach stability in his B cutoff point.

It is interesting to note that, if conditional probabilities are disregarded, the proportion of A and B predictions closely

match their simple probabilities of occurrence (the last points in Fig. 6—78:22 vs. 76:24). Such data might lead to the erroneous conclusion that *S* completely ignores stimulus readings and does, in fact, match the probabilities of reinforcement associated with the two meters. The fallaciousness of this interpretation is clearly shown in the empirical cutoff data; for *S* to choose Meter A, Stimulus A must, on the average, be above 25.6 or Stimulus B below 20.6. Were he to ignore the readings completely, there would be no reliable cutoff points at all, and were he to consider the two sets of readings as drawn from a common population, a single cutoff point would appear. Obviously, his predictions are influenced both by the magnitude of the readings and by the distribution from which they are drawn.

A final point of interest is found in the data for judgment of conditional probabilities (Fig. 5). Median judgments are most accurate for probabilities below .50; higher probabilities are consistently underestimated. Improvement over trials appears largely in a reduction of this underestimation. In spite of the fact that B > A occurred much more frequently than A > B in the prediction trials, judgments of the two kinds of events show surprisingly similar trends over the full range of conditional probabilities.

REFERENCES

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

Edwards, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Estes, W. K., Burke, C. J., Atkinson, R. C., & Frankmann, J. P. Probabilistic discrimination learning. *J. exp. Psychol.*, 1957, **54**, 233–239.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1951, **42**, 1–5.

Neimark, E. D., & Shuford, E. H. Comparison of predictions and estimates in a probability learning situation. *J. exp. Psychol.*, 1959, **57**, 294–298.

Restle, F. *Psychology of judgment and choice: A theoretical essay.* New York: Wiley, 1961.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

Siegel, S., & Goldstein, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 293–340.

(Received December 19, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 100–101

SUPPLEMENTARY REPORTS

FUNCTIONAL STIMULUS LEARNING AS RELATED TO DEGREE OF PRACTICE AND MEANINGFULNESS[1]

KENNETH L. LEICHT AND DONALD H. KAUSLER

Saint Louis University

Functional stimulus learning for paired nonsense syllables, as measured by recognition of stimuli paired with specific responses, varied directly with both stimulus meaningfulness and degree of S-R practice, with scores exceeding S-R anticipations until a common asymptote was attained. Further, letter identification indicated that more letters of the stimulus were utilized as the functional stimulus under high than low meaningfulness, suggesting a convergence between the nominal and functional stimulus as meaningfulness increases.

Stimulus learning during paired-associate practice has ordinarily been measured by R-S recall. In terms of the distinction between the nominal and functional stimulus (Underwood, 1963), such measures reflect the amount of nominal stimulus learning occurring during S-R acquisition. However, as Underwood and Schulz (1960) have indicated, with nonsense syllables as stimuli, a single element may serve as the functional stimulus. For such terms, recall of the remaining elements may represent an incidentally learned integration beyond that required for establishing discernible cues.

R-S recall for nonsense syllables has been demonstrated (Jantz & Underwood, 1958) to increase with S-R practice until an asymptote is reached which is considerably below the S-R asymptote. Similarly, a monotonic relationship has been found (Cassem & Kausler, 1962) between stimulus meaningfulness (M) and amount of R-S recall. This experiment extended these two variables to the functional stimulus, with stimulus learning being measured by both a recognition and a letter identification procedure.

Method.—A 4 $\times$ 2 mixed design was employed, with S-R practice (6, 12, 18, or 30 trials) as "between" and stimulus M—high (H) or low (L)—as "within" variables. There were 14 undergraduates in each of the four practice groups. For the 6-, 12-, and 18-trial groups, *S* received that number of trials whether or not a criterion of one perfect trial had been attained. The number of *S*s reaching criterion was 0, 0, and 2 for 6, 12, and 18 trials, respectively. In order to avoid overlearning, *S*s in the remaining group received 30 trials or trials to criterion, whichever came first (8 *S*s reached criterion; mean number trials = 23.14). The list was mixed, with 5 HM (mean Archer, 1960, value = 83.5%) and 5 LM (mean = 12.8%) stimulus terms. For control purposes, two combinations of the list were prepared, differing in specific S-R pairings, with three different serial orders for each list. Intralist similarity was made as low as possible. Practice was at a 2:2-sec. rate, with a 4-sec. intertrial interval.

A recognition trial followed S-R practice, in which *S* was presented cards, one at a time, containing responses on the left side and three alternative stimuli on the right. One of the stimuli was correct, the other two being stimuli associated with other responses in the list (one LM and one HM). The *S* was instructed to encircle the correct stimulus. For letter identification, *S* was given a mimeographed sheet, following the recognition trial, containing all S-R pairings, and asked to encircle the letters of the stimulus utilized in associating the response to that stimulus.

Results.—Mean recognition scores, corrected for guessing, are given in Table 1, together with summary S-R data. An analysis of variance revealed a significant effect on recognition for Trials, $F\,(3, 52) = 10.04$, $p < .001$, and Meaningfulness, $F\,(1, 52) = 27.30$, $p < .001$. Although the Trials $\times$ M interaction suggested a trend, with HM scores reaching an asymptote relatively early in practice and LM later, the effect was not significant, $F\,(3, 52) = 1.96$, $p > .10$. As would be expected, functional stimulus learning progresses more rapidly than S-R learning (see Table 1), with the latter requiring additional processes. Further, as is to be expected, functional stimulus learning, unlike R-S recall, reaches the same asymptote as S-R acquisition. That is, complete S-R

[1] This study is based on a thesis submitted by the first author to the Graduate School, St. Louis University in partial fulfillment for the Master of Arts degree.

TABLE 1

SUMMARY DATA FOR STIMULUS RECOGNITION AND S-R ACQUISITION

Trials	Stimulus Recognition Meaningfulness				S-R Acquisition Meaningfulness			
	HM		LM		HM		LM	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
6	2.86	1.63	2.00	1.32	2.07	1.49	1.21	.89
12	4.46	.79	2.75	1.86	3.21	1.48	2.64	1.74
18	4.46	.79	2.86	1.74	4.28	.82	2.64	1.45
30	4.68	.20	4.25	1.28	4.57	.94	4.36	.93

learning is impossible without the prior establishment of reliable functional stimulus cues.

Mean letter identification scores (maximum = 15) for the 6-, 12-, 18-, and 30-trial groups, respectively, were 11.36, 11.64, 12.28, and 12.64 for HM and 7.43, 8.57, 8.21, and 9.57 for LM. An analysis of variance indicated a pronounced effect for M, F (1, 52) = 46.24, $p < .001$, but the Trials and Trials × M effects fell far short of significance (Fs <1). These results suggest that there may be an increasing convergence between the nominal and functional stimulus as M increases.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, 74(10, Whole No. 497).

CASSEM, N., & KAUSLER, D. H. Supplementary report: Effects of stimulus association value and exposure duration on R-S learning. *J. exp. Psychol.*, 1962, 64, 94.

JANTZ, E. M., & UNDERWOOD, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, 56, 174–179.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer (Ed.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received December 2, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 101–103

UCS INTENSITY AND THE ASSOCIATIVE STRENGTH OF THE EYELID CR WITH A MASKED CONDITIONING PROCEDURE [1]

MARVIN J. HOMZIE [2] AND GERALD WEISS

State University of Iowa

60 human Ss were given 50 eyelid conditioning trials while engaged in a probability-learning task introduced for the purpose of masking the conditioning experiment. For the high-reinforcement group (H′) on half of the trials a strong 2.0-psi air puff occurred 500 msec. after the onset of the 500-cps CS; on the other half of the trials a weak .33-psi air puff was presented alone without the CS. For the low-reinforcement group (L′) these conditions were reversed. Since the average UCS intensity was identical for the 2 groups, their drive levels (D) presumably were equated. The H′ group reached a higher performance level than the L′ group (p < .005). These results were in agreement with a previous study which employed more "standard" conditioning procedures. It was concluded that habit strength (H) of a defense CR is a function of the intensity of the UCS.

In an earlier study reported by Spence, Haggard, and Ross (1958) it was found that a group of *S*s that had a strong UCS on the half of the trials that involved pairing of the CS-UCS and a weak UCS alone on the other half of the trials gave a significantly higher number of CRs than a group in which the strengths of the UCS on the two types of trials were reversed. Because level of drive (D) in the two groups was equated this finding was interpreted as lending support to

[1] This study was carried out as part of a project concerned with the influence of motivation on performance in learning under Contract Nonr-1509(04), Project NR 154-107 between the State University of Iowa and the Office of Naval Research. The project is under the direction of Kenneth W. Spence.

The first author was a United States Public Health Service Predoctoral Fellow MPM-16,385 during the period of this investigation.

[2] Now at the University of Denver.

a reinforcement-type interpretation that habit strength (H) in aversive conditioning is a function of the intensity of the UCS. The present experiment replicates the above study with the conditioning embedded (masked) within the context of a probability-learning task. The advantage of this procedure is that it attenuates the effects of human cognitive or set factors (Spence, 1963; Spence, Homzie, & Rutledge, 1964). The purpose of this study was to determine if this procedure would produce data consistent with previously obtained results. Additional procedural differences between the two studies are: the absence of a ready signal, 500-cps tone as the CS instead of a light, and a shorter intertrial interval.

Method.—A total of 60 men and women students from an introductory course in psychology formed the two experimental groups, each of which contained 15 men and 15 women. The apparatus was identical to that of Spence, Homzie, and Rutledge (1964). A CR was recorded whenever the Brush recording pen showed a deflection of 1 mm. or more in the range 150–500 msec. following the onset of the CS.

For the high-reinforcement group (H′) a strong puff (2.0 psi) always occurred on a conditioning trial, whereas a weak puff (.33 psi) was presented alone without the CS. For the low-reinforcement group (L′) these conditions were reversed, the weak puff being presented on the conditioning trials and the strong puff being presented as an extra stimulus. The order of presentation of conditioning and extrastimulus trials was prearranged according to an irregular order in which the number of each was equalized in blocks of four trials.

The *S*s were given a set of instructions to the effect that the experiment was concerned with the effects of distraction on performance in a difficult problem-solving situation. They were told that their task was to predict, on the onset of a central signal light, which of two small lamps located to the left and right of the signal light was going to light up and to signify their prediction by pressing one of two push buttons located on the left or right arm of their chair. They were told that distracting stimuli in the form of an air puff to their eye and a tone would be given between their response of pressing a button and the lighting up of one of the small lamps. The onset of the CS (500-cps tone) came 2 sec. after the onset of the signal light and had a duration of 2,550 msec. The duration of the signal light was 4,550 msec. On conditioning trials the onset of the UCS (puff) followed the onset of the CS by 500 msec. and lasted for 50 msec. On extrastimulus (UCS alone) trials the UCS was presented 4,500 msec. after the onset of the signal light. Five hundred milliseconds after the termination of the signal light one or the other of the small lights came on and remained on for 1 sec. The intertrial period was at predetermined intervals of 9, 12, and 15 sec., averaging 12 sec. for each *S*. Fifty conditioning (i.e., paired CS-UCS) trials were given.

Results and discussion.—Figure 1 presents the frequency curves of conditioning for the high-reinforcement group (H′) and the low-reinforcement group (L′). Also included in Fig. 1 are the performance curves from a study previously reported by Spence, Haggard, and Ross (1958) for *S*s conditioned under "standard" conditions, i.e., without the probability-learning task. The curves labeled Gp. H and Gp. L are the conditioning curves for the high- and low-reinforcement groups, respectively. As may be seen, in both experiments, the curve for the high-reinforcement group (H′, H) is well above that for the low-reinforcement group (L′, L). In agreement with the previous findings, evaluation of the performance levels over the last 20 conditioning trials by means of the Mann-Whitney U test resulted in a significant difference between Groups H′ and L′ ($p < .005$).

In view of the procedural differences, the results of the two experiments are remarkably similar. Thus, conditioning embedded in the context of the probability-learning task produces results consistent with those previously established under more standard condi-

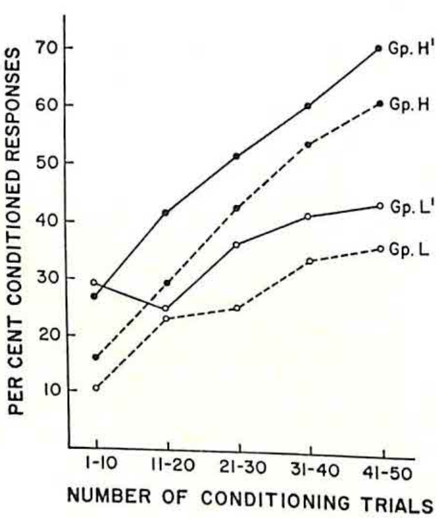


FIG. 1. Acquisition curves showing the percentage of CRs in successive blocks of 10 conditioning trials.

tions. These two experiments indicate that with level of drive (D) equated, performance in classical aversive conditioning is a function of the intensity of the UCS occurring on the reinforced trials. These results may be interpreted as supporting a reinforcement-type theory that habit strength (H) in classical aversive conditioning is a function of the intensity of the UCS occurring on the reinforced trials.

REFERENCES

SPENCE, K. W. Cognitive factors in the extinction of the conditioned eyelid response in humans. *Science*, 1963, 140, 1224–1225.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, 55, 404–411.

SPENCE, K. W., HOMZIE, M. J., & RUTLEDGE, E.F. Extinction of the human eyelid CR as a function of the discriminability of the change from acquisition to extinction. *J. exp. Psychol.*, 1964, 67, 545–552.

(Received November 22, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 103–105

PREDICTION OF FREE RECALL FROM WORD-ASSOCIATION MEASURES: A REPLICATION

ARTHUR M. BODIN, LEWIS A. CRAPSI, MARILYN R. DEAK, THEOBOLD R. MORDAY, AND LAURENCE D. RUST

State University of New York at Buffalo

Mean recall frequency of 99 Kent-Rosanoff words was investigated as a function of cue number, a measure of intralist association. Word cards were shuffled and viewed for 5 sec. each by 15 female and 13 male undergraduate Ss, 17–19 yr. old. Following a 15-min. rest, Ss had 5 min. for written recall. Results showed a significant relationship between cue number (Nc) and mean recall frequency. Peaks and valleys graphed from replication and original data matched and coincided with Nc's unique to one word, suggesting these fluctuations reflect particular, not random effects. Cue linkage increased with Nc and depended on association strength between cuing and cue-linked words, replicating original findings, and, originally unreported, favored early recall.

This study was an attempt to replicate Rothkopf and Coke's (1961b) investigation of the relationship between the free recall of each word in a word-association list and *cue number* (Nc). As computed by Rothkopf and Coke (1961a), Nc is the number of other words in the list which elicited a particular word more than 11 times *or* as one of the 10 most frequent associates in the 1952 Minnesota norms of the Kent-Rosanoff Word-Association Test (Russell & Jenkins, 1954). Thus, Nc quantifies the degree of intralist associations, which Deese (1959, 1960) showed to be positively related to the number of items recalled.

Method.—The 100-word Kent-Rosanoff list was modified by the deletion of CHEESE, leaving 99 words and permitting equal division into three sublists for presentation in turn to three *S*s per sitting. Each word was typed in capitals on a separate 3 × 5 in. white card.

Twenty-eight undergraduate students at the State University of New York at Buffalo served as *S*s without pay. These 15 females and 13 males ranged from 17 to 19 yr. of age.

The stimulus card deck was shuffled and split into three numerically equal subdecks, placed, one before each *S*, face up, but covered with a blank top card.

Instructions were read aloud as follows:

> This experiment will take about half an hour. You will hold a stack of 3 × 5 cards. On each card there is a different word, except for the card on top, which is blank. You inspect each card for a few seconds, until I say, "Next"; then turn it over onto the table and inspect the next word. Once we start, please don't talk till we're finished with the whole experiment. I'll tell you when that is. Do you have any questions? —All right: when I say, "Begin," turn the top card over and start.

Subdecks were inspected at the rate of 5 sec. per card, then shuffled separately by *E* and reassigned, until each subdeck had been completed by each *S*, in rotation. A 15-min. rest pause was introduced as follows:

> Now please fill this out. [The *E* handed each *S* a biographical data form.] Then take a rest pause in the hall right outside

> this door till we call you back for the next in about 10 minutes; but please do not discuss this experiment.

After the interlude, *S*s were given 5 min. to recall the stimulus words and were instructed as follows:

> When I say, "Begin," write down all the words you can remember from the cards you looked at earlier. Write each word on a separate page. Do not tear off any pages. Write as many words as you can. I will stop you in 5 minutes. Begin.

Results and discussion.—Results have been plotted in Fig. 1 as mean recall frequency of words having the same Nc vs. Nc, beneath which, in parentheses, is the number of words in the deck having that Nc. It will be noted that the peaks and valleys obtained in this replication are consistently congruent with those obtained by Rothkopf and Coke (1961b). And the fluctuations in mean recall frequency, which seemed in the original data to reflect mere random variations, now appear to be functions of the particular words involved. This impression is strengthened by the fact that the peaks and valleys, with only one exception, occur in conjunction with an Nc to which only one word contributed data. Thus, no opportunity for idiosyncrasies to be averaged out was afforded by these instances, that is, they precluded possibilities for other words to smooth or balance out any recall effects peculiar to particular words of a given Nc. Restating this evidence, every Nc which was unshared, that is to say, unique to one word, coincides with either a peak or a valley in recall frequency.

The consistently higher mean recall frequencies obtained by Rothkopf and Coke (1961b) may be a function of a difference in set. The original unpublished instructions included the statement, "Look at the words carefully because after you have seen all the words, I am going to ask you to remember as

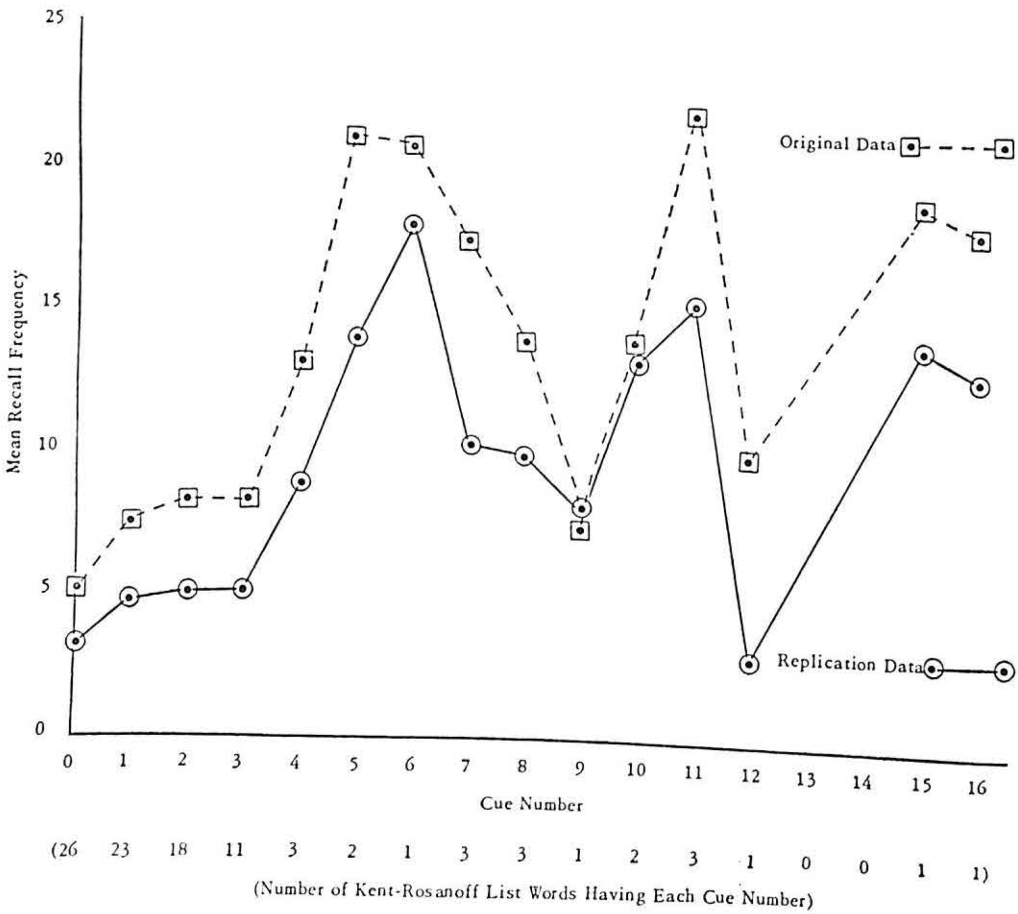


FIG. 1. Mean recall frequencies for words of each cue number.

many of them as you can."[1] The replication instructions, however, made no reference to a future recall task. Since *S*s in the two experiments may have differed in relevant ways other than the set induced in them, this interpretation of a consistent difference in results as a function of a difference in instructions cannot be advanced as more than a possibility, attractive because of its face validity.

The product-moment correlation coefficient between recall frequency and Nc was .40 ($p < .001$), a result in essential agreement with that of the original study ($r = .62$, $p < .001$).

An analysis of variance showed that variance between Nc was significantly greater than within-Nc variance, $F\ (14, 84) = 9.02$, $p < .001$. Again, the replication result agrees closely with the result of the original study, $F\ (14, 84) = 7.81$, $p < .001$.

Rothkopf and Coke (1961b) suggested that the effect of Nc on recall might be accounted for by a cuing mechanism having greater strength at higher Nc levels. To measure such a cuing mechanism, a tally was made of *cue linkage*, i.e., the number of instances in which each word was recalled immediately after a word for which it was an association response. The cue-linkage frequency of each word was divided by the corresponding number of opportunities for cue linkage, the latter being each word's frequency of recall in any ordinal position except the first one in a list. The ratio of cue linkage to opportunities for cue linkage appeared to be an increasing function of Nc. The analysis of variance relevant to the present data showed that variance between Nc was significantly greater than within-Nc variance, $F\ (13, 59) = 2.07$, $p < .05$. Again, this result confirms the original findings, $F\ (13, 59) = 5.60$, $p < .001$.

To find whether cue linkage depended on association strength between cuing and cue-linked words, a three-way tally was made, according to whether the association frequency of the latter was greater than, equal to, or less than the median association frequency of the relevant normative associations to the cuing words as computed by Rothkopf and Coke (1961a). Of 124 cue linkages in 675 recalled words, 81 were above the median, 16 were equal to the median, and 27 were below the median. The 16 cases at the median were then split equally into the above and below median categories, and a binomial test was performed on the data (Siegel, 1956, pp. 36–42). The two-tailed test was significant ($z = 4.76$, $p < .001$). Again, the present results support the findings of the original study ($z = 7.31$, $p < .001$). Rothkopf and Coke (1961b) concluded:

> Words which followed one of their cues in recall tended to be among those word associations which had association frequencies above the median association frequency of the association responses (list words only) to the cue word [p. 438].

One additional analysis, not reported by Rothkopf and Coke (1961b) was performed in the present study. A count was made of the number of cue linkages which occurred in the first and second halves of each *S*'s recall list, excluding the initial word, since it could not manifest cue linkage in the absence of a cue. The hypothesis was that more cue linkages would be found among the words written in the first than in the second half of the recall lists. This proved to be true for 19 of the 28 *S*s. Of the 9 remaining *S*s, 6 produced lists having an equal number of cue linkages in each half, and only 3 produced lists with more cue linkages in the second half. Thus, the hypothesis was confirmed by a one-tailed sign test which showed that the obtained difference in the predicted direction was significant ($p < .001$).

The summarized conclusion of Rothkopf and Coke (1961b) was that "These results were consistent with a cuing interpretation of the frequency and sequence of recall phenomena which were the subject of this study [p. 438]." Every aspect of the present replication supports their conclusion.

[1] E. Z. Rothkopf, personal communication, February 26, 1962.

REFERENCES

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

DEESE, J. Frequency of usage and number of words in free recall: The role of association. *Psychol. Rep.*, 1960, **7**, 337–344.

ROTHKOPF, E. Z., & COKE, E. U. Intralist association data for 99 words of the Kent-Rosanoff word list. *Psychol. Rep.*, 1961, **8**, 463–474. (a)

ROTHKOPF, E. Z., & COKE, E. U. The prediction of free recall from word association measures. *J. exp. Psychol.*, 1961, **62**, 433–438. (b)

RUSSELL, W. A., & JENKINS, J. J. Language in behavior: The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* N. Y.: McGraw-Hill, 1956.

(Received November 29, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 1, 106–110

Since 1962 APA's Project on Scientific Information Exchange in Psychology has conducted a series of studies which have led to a detailed description of the dissemination of scientific information in psychology. It seemed clear from this description that one of the most important and difficult communication activities of the scientist is his effort to establish and maintain contact with current work relevant to his interests. Assistance to the active user of scientific information in such activities seems the most neglected area in the development of innovations to improve scientific information exchange. During the next 12–15 months the Project will sponsor a series of innovations which have been designed to enhance scientific communication among psychologists. These innovations have been planned as limited trials, wholly supported by the Project for the purpose of carrying out studies on the effects of such innovations on scientific information exchange.

The Editor of the *Journal of Experimental Psychology* and his staff have agreed to cooperate in testing one of the innovations. Beginning with this issue and continuing through the remainder of Volumes 69 and 70 (1965), the titles and authors of accepted papers will be listed following Supplementary Reports. Such listing will make it possible for readers to become aware of research many months in the advance of its journal publication. At the end of the year's trial the results will be evaluated and further consideration given to the advisability of continuing this service.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Function Order and Paired-Associate Learning: Cameron R. Peterson,* Z. J. Ulehla, and Richard S. Lehman: Engineering Psychology Laboratory, Institute of Science and Technology, Box 618, University of Michigan, Ann Arbor, Michigan.

Rate of Presentation in Serial Learning: Geoffrey Keppel* and Robert J. Rehula: Department of Psychology, University of California, Berkeley, California 94720.

Leadership in Small Groups: A Mathematical Approach: Arnold Binder,* Burton R. Wolin, and Stanley J. Terebinski: Department of Psychology, Indiana University, Bloomington, Indiana.

Effects of Differential Value on Recall of Visual Symbols: Harvey A. Taub*: Cornell Aeronautical Laboratory, Inc., P.O. Box 325, Buffalo, New York.

Concept Learning and Verbal Control under Partial Reinforcement and Subsequent Reversal or Nonreversal Shifts: Daniel C. O'Connell*: 221 North Grand Boulevard, St. Louis, Missouri.

Effect of Intertrial Activity on the Relationship between Awareness and Verbal Operant Conditioning: Paul W. Dixon and William F. Oakes*: Department of Psychology, University of Hawaii, Honolulu 14, Hawaii.

The Partial-Reinforcement Effect Sustained through Blocks of Continuous Reinforcement in Classical Eyelid Conditioning: Sally L. Perry and John W. Moore*: Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Bases for Preferences among Three-Outcome Bets: Sarah Lichtenstein*: Department of Psychology, Southern Illinois University, Carbondale, Illinois.

Apparent Spatial Position and the Perception of Lightness: Jacob Beck*: Psychological Laboratories, Memorial Hall, Harvard University, Cambridge, Massachusetts 02138.

A Test of the "Limited Capacity" Hypothesis: Bennet B. Murdock, Jr.*: Department of Psychology, Waterman Building, University of Vermont, Burlington, Vermont.

"Fate" of List 1 R-S Associations in Transfer Theory: Norma R. Ellington and Donald H. Kausler*: Department of Psychology, Saint Louis University, 221 North Grand Boulevard, Saint Louis 3, Missouri.

Facilitation of Competing Responses as a Function of "Subnormal" Drive Conditions: Charles Y. Nakamura* and William E. Broen, Jr.: Department of Psychology, University of California, Los Angeles, California 90024.

---

* Asterisk indicates author for whom the address is supplied. Manuscripts are listed in order of acceptance; the first 15 will probably appear next month.

Similarity in Stimulus Material and Stimulus Task on the Formation of a New Scale of Judgment: M. E. Tresselt*: Department of Psychology, New York University, Washington Square, New York, New York 10003.

Unlearning as a Function of the Relationship between Successive Response Classes: Leo Postman, Geoffrey Keppel,* and Karen Stark: Department of Psychology, University of California, Berkeley, California 94720.

UCS Properties in Classical Conditioning of the Albino Rabbit's Nictitating Membrane Response: Alfred Bruner*: Department of Physiology, School of Medicine, University of California Medical Center, Los Angeles, California 90024.

Monoptic and Dichoptic Visual Masking by Patterns and Flashes: Peter H. Schiller*: Psychology Section, E10-137, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Application of a Markov Model to Free Recall and Recognition: Walter Kintsch* and Charles J. Morris: Department of Psychology, University of Missouri, Columbia, Missouri.

Cultural Primaries as a Source of Interference in Short-Term Verbal Retention: Kenneth A. Blick*: Department of Psychology, Randolph-Macon College, Ashland, Virginia.

Some Effects of Contour on Simultaneous Brightness Contrast: Phyllis W. Berman* and H. W. Leibowitz: 1017 23rd Avenue S.E., Minneapolis, Minnesota.

Preexposure to Visually Presented Forms and Nondifferential Reinforcement in Perceptual Learning: Larry C. Kerpelman*: Division of Cognitive Development, Institute for Juvenile Research, 608 South Dearborn Street, Chicago, Illinois 60605.

Differential Visual Feedback of Component Motions: John D. Gould*: IBM Watson Research Center, P.O. Box 218, Yorktown Heights, New York.

Training and Timing in the Generalization of a Voluntary Response: Sheldon H. White*: Center for Cognitive Studies, Harvard University, Cambridge, Massachusetts 02138.

Effect of Prior Knowledge of the Stimulus on Word-Recognition Processes: Ralph Norman Haber*: Department of Psychology, University of Rochester, Rochester, New York.

Effects of Stimulus Change upon the GSR and Reaction Time: Paul F. Grim and Sheldon H. White*: Center for Cognitive Studies, Harvard University, Cambridge, Massachusetts 02138.

Further Effects of Subject-Generated Recoding Cues on Short-Term Memory: Richard H. Lindley* and Shari E. Nedler: Department of Psychology, Eastern Michigan University, Ypsilanti, Michigan.

Effect of CS-Onset UCS-Termination Delay, UCS Duration, CS-Onset UCS-Onset Interval, and Number of CS-UCS Pairings on Conditioned Fear Response: Andrew Strouthes*: Division of Science and Mathematics, Harpur College, Binghamton, New York 13901.

Perception of Off-Size Versions of a Familiar Object under Conditions of Rich Information: Samuel Fillenbaum,* H. Richard Schiffman, and James Butcher: Department of Psychology, University of North Carolina, Chapel Hill, North Carolina.

Spontaneous Recovery of Letter-Sequence Habits: Eli Saltz*: Center for the Study of Cognitive Processes, Wayne State University, Detroit, Michigan 48202.

Moments of Area and of the Perimeter of Visual Form as Predictors of Discrimination Performance: Leonard Zusne*: Department of Psychology, University of Tulsa, Tulsa, Oklahoma 74104.

Effects of Drive, Reinforcement Schedule, and Change of Schedule on Performance: Pietro Badia*: Department of Psychology, Bowling Green State University, Bowling Green, Ohio.

First-List Retention as a Function of the Method of Recall: John P. Houston,* Bertram E. Garskof, Dale E. Noyd, and Janice M. Erskine: Department of Psychology, University of California, Los Angeles, California 90024.

Associative and Differentiation Variables in All-or-None Learning: Clessen J. Martin*: 441 Erickson Hall, Michigan State University, East Lansing, Michigan.

Event Salience and Response Frequency in a Ten-Alternative Probability-Learning Situation: Lee Roy Beach* and Richard W. Schoenberger: Department of the Navy, Office of Naval Research, Washington, D. C. 20025.

Cue and Secondary Reinforcement Effects with Children: Joseph B. Sidowski,* Norman Kass, and Helen Wilson: Department of Psychology, San Diego State College, San Diego, California 92115.

Temporal Determinants of a Kinesthetic Aftereffect: G. Singer* and R. H. Day: Department of Psychology, University of Sydney, Sydney, N.S.W. Australia.

Effect of Extraneous Stimulation on the Visual Perception of Verticality: A Failure to Replicate: Robert Fried and Richard G. Lathrop*: Division of Education and Psychology, Chico State College, Chico, California 95927.

Effects of Shock Intensity and Placement on the Learning of a Food-Reinforced Brightness Discrimination: Elizabeth B. Curlin and John W. Donahoe*: Center for Brain Research, University of Rochester, River Campus Station, Rochester 20, New York.

Improvement of Visual and Tactual Form Discrimination: Ann D. Pick*: Department of Psychology, Macalester College, Saint Paul 1, Minnesota.

Effects of CS and UCS Change on Extinction of the Conditioned Eyelid Response: Louis E. Price,* David W. Abbott, and William E. Vandament: Department of Psychology, University of Massachusetts, Amherst, Massachusetts.

Stimulus Generalization as a Function of Drive Level and the Relation between Two Measures of Response Strength: J. Robert Newman and G. Robert Grice*: University of Illinois, Department of Psychology, Urbana, Illinois 61803.

Differential Effects of Stimulus and Response Isolation in Paired-Associate Learning: Raymond L. Erickson*: Department of Psychology, University of New Hampshire, Durham, New Hampshire 03824.

Subjective Intensity of Mineral Taste in Water: William H. Bruvold* and William R. Gaffey: University of California, School of Public Health, Earl Warren Hall, Berkeley, California 94720.

Stimulus Generalization as a Function of Drive Shift: Robert B. Zajonc and David V. Cross*: Department of Psychology, University of Michigan, Ann Arbor, Michigan.

Overlearning and Brightness-Discrimination Reversal: M. R. D'Amato* and Donald Schiff: Department of Psychology, New York University, University Heights, New York, New York 10053.

Inferred Components of Reaction Times as Functions of Foreperiod Duration: Raymond H. Hohle*: Institute of Child Behavior and Development, State University of Iowa, Iowa City, Iowa 52240.

First-List Retention as a Function of List Differentiation and Second-List Massed and Distributed Practice: John P. Houston* and James H. Reynolds: Department of Psychology, University of California, Los Angeles, California 90024.

Effect of Luminance, Exposure Duration, and Task Complexity on Reaction Time: Jaques Kaswan* and Stephen Young: Department of Psychology, University of California, Los Angeles, California 90024.

Signal Uncertainty and Sleep Loss: Harold L. Williams,* Ometta F. Kearney, and Ardie Lubin: Department of Psychiatry, Neurology, and Behavioral Sciences, University of Oklahoma School of Medicine, 800 N.E. 13th St., Oklahoma City 4, Oklahoma.

S-R Compatibility and Information Reduction: Paul M. Fitts* and Irving Biederman: Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

Distribution and Sequence Effects in Judgment: Allen Parducci* and Arthur Sandusky: Department of Psychology, University of California, Los Angeles, California 90024.

Effects of Delayed Auditory Feedback on Morse Transmission by Skilled Operators: Aubrey J. Yates*: Department of Psychology, University of Western Australia, Perth, Australia.

Stimulus Durations and Total Learning Time in Paired-Associates Learning: Cavlin F. Nodine*: Department of Psychology, Miami University, Oxford, Ohio.

Mediation Instructions versus Unlearning Instructions in the A-B, A-C Paradigm: Kent M. Dallett* and Lester D'Andrea: Department of Psychology, University of California, Los Angeles, California 90024.

Compound Stimuli, Drive Strength, and Primary Stimulus Generalization: Albert F. Healey*: Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut.

Discrimination Performance as Affected by Problem Difficulty and Shock for Either the Correct or Incorrect Response: Harry Fowler and George J. Wischner*: Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

Delayed Cold-Induced Vasodilation and Behavior: Warren H. Teichner*: Institute of Environmental Psychophysiology, Clark Hall, University of Massachusetts, Amherst, Massachusetts.

The Reinforcement Relation as a Function of Instrumental Response Base Rate: Robert W. Schaeffer*: Department of Psychology, Florida State University, Tallahassee, Florida.

A Failure to Find a Response Persisting in the Apparent Absence of Motivation: Milton A. Trapold,* Sandra R. Blehert, and Thomas Sturm: Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

Meaningfulness as a Variable in Dichotic Hearing: David S. Emmerich,* Donald M. Goldenbaum, Dale L. Hayden, Linda S. Hoffman, and Jeanne L. Treffts: Department of Psychology, Indiana University, Bloomington, Indiana.

Effect of Pairing Directionality and Anticipatory Cue in Paired-Associate Learning: James F. Voss*: Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania.

Abolition of the PRE by Instructions in GSR Conditioning: Wagner H. Bridger* and Irwin J. Mandel: Department of Psychiatry, Albert Einstein College of Medicine, Yeshiva University, Eastchester Road and Morris Park Avenue, New York, New York 10461.

Strength of the Relationship between the Value of an Event and its Subjective Probability as a Function of Method of Measurement: Dean G. Pruitt* and Robert D. Hoge: Center for Research on Social Behavior, University of Delaware, Newark, Delaware.

Loudness, a Product of Volume Times Density: S. S. Stevens,* Miguelina Guirao, and A. Wayne Slawson: Psycho-Acoustic Laboratory, Harvard University, Cambridge, Massachusetts 02138.

Prior Context and Fractional versus Multiple Estimates of the Reflectance of Grays against a Fixed Standard: E. C. Poulton* and D. C. V. Simmonds: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

The Role of Apparent Slant in Shape Judgments: Wilma A. Winnick* and Ilana Rogoff: Department of Psychology, Queens College, Flushing 67, New York.

Stimulus Determinants of Choice Behavior in Visual Pattern Discrimination: Jaques Kaswan,* Stephen Young, and Charles Y. Nakamura: Department of Psychology, University of California, Los Angeles, California 90024.

Effect of Stimulus Variables on Choice Reaction Times and Thresholds: Jaques Kaswan* and Stephen Young: Department of Psychology, University of California, Los Angeles, California 90024.

Resistance to Extinction Following Blocking of the Instrumental Response during Acquisition: W. Edward Bacon*: Department of Psychology, McGill University, Montreal, Canada.

Neutralization of Stimulus Bias in the Rating of Grays: Irwin Pollack*: Mental Health Research Institute, University of Michigan, Ann Arbor, Michigan 48104.

Reward versus Nonreward in a Simultaneous Discrimination: R. Allen Gardner* and W. B. Coate: Department of Psychology, University of Nevada, Reno, Nevada.

Sample Size and the Revision of Subjective Probabilities: Cameron R. Peterson,* Robert J. Schneider, and Alan J. Miller: Engineering Psychology Laboratory, Institute of Science and Technology, Box 618, University of Michigan, Ann Arbor, Michigan.

Effects of Postresponse Stimulus Duration upon Discrimination Learning in Human Subjects: Donald J. Dickerson* and Norman R. Ellis: George Peabody College for Teachers, Nashville 5, Tennessee.

Recognition Memory for Random Shapes as a Function of Complexity, Association Value, and Delay: Herbert James Clark*: 6814 Hubbard Drive, Dayton 24, Ohio.

Parameters of Paired-Associate Verbal Learning: Length of List, Meaningfulness, Rate of Presentation, and Ability: John B. Carroll* and Mary Long Burke: Laboratory for Research in Instruction, Graduate School of Education, Harvard University, Longfellow Hall, Appian Way, Cambridge, Massachusetts 02138.

Reaction Time to Changes in the Intensity of White Noise: David H. Raab* and Mitchell Grossberg: Department of Psychology, Brooklyn College, Brooklyn, New York 10010.

Habituation of Alternation Behavior: P. E. Freedman*: Psychology Laboratories, Pioneering Research Division, U. S. Army Natick Laboratories, Natick, Massachusetts.

Discriminability and Preference for Attributes in Free and Constrained Classification: Shiro Imai and W. R. Garner*: Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218.

Children's Performance as a Function of the Degree of Visual Stimulus Deprivation and Satiation: Richard D. Odom*: Institute of Child Development, University of Minnesota, Minneapolis 14, Minnesota.

Concept Identification: Effects of Varying Length and Informational Components of the Intertrial Interval: Lyle E. Bourne, Jr.,* Donald E. Guy, David H. Dodd, and Don R. Justesen: Department of Psychology, University of Colorado, Boulder, Colorado 80304.

Primacy and Recency Effects in Serial-Position Curves of Immediate Recall: John C. Jahnke*: 409 Berkely Street, Ann Arbor, Michigan.

Paired-Associates Learning with Varying Relative Percentages of Occurrence of Alternative Response Members: Influence of Instructions: Albert E. Goss*: Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Coherence Detection: A Basic Mechanism: Albert S. Rodwan*: Department of Psychology, Emory University, Atlanta, Georgia 30322.

Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception: D. V. Cross, H. L. Lane,* and W. C. Sheppard: Behavior Analysis Laboratory, 1315 Hill Street, University of Michigan, Ann Arbor, Michigan 48104.

Scaling Musical Preferences on the Prothetic Nature of Affective Judgments: Soon Duk Koh*: Hastings State Hospital, Hastings, Minnesota.

Short-Term Retention of Verbal Units with Equated Degrees of Learning: John P. Houston*: Department of Psychology, University of California, Los Angeles, California 90024.

Effects of Ordered and Constant Sucrose Concentrations on Nonreinforced Performance: Tom N. Tombaugh* and Melvin H. Marx: Department of Psychology, University of Missouri, 17 McAlester Hall, Columbia, Missouri 65202.

Vigilance Performance with a Qualitative Shift in Reinforcers: William Bevan* and Edward D. Turner: Department of Psychology, Kansas State University, Manhattan, Kansas.

Relationship between Static and Dynamic Visual Acuity: Seymour Weissman* and C. M. Freeburne: U. S. Naval Medical Research Laboratory, U. S. Naval Submarine Base New London, Groton, Connecticut 06342.

Effect of Successive Addition of Stimulus Elements on Paired-Associate Learning: Sam C. Brown, William F. Battig,* and Richard Pearlstein: Department of Psychology, University of Maryland, College Park, Maryland.

Steepness of Approach and Avoidance Gradients in Humans as a Function of Experience: Theory and Experiment: Seymour Epstein* and Walter D. Fenz: Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Sources of Transfer from Original Training to Discrimination Reversal: W. B. Coate and R. Allen Gardner*: Department of Psychology, University of Nevada, Reno, Nevada.

A Test of a Discrimination Hypothesis of Secondary Reinforcement: Nancy A. Myers and Jerome L. Myers*: Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Acoustic Similarity and Intrusion Errors in Short-Term Memory: Wayne A. Wickelgren*: Psychology Section, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

The Relation of Drive to Finger-Withdrawal Conditioning: Merrill F. Elias*: R. D. 5, Rome, New York 13440.

False Recognition Produced by Implicit Verbal Responses: Benton J. Underwood*: Department of Psychology, Northwestern University, Evanston, Illinois 60201.

Effect of UCS Intensity upon the Acquisition of Conditioned Responses Acquired under a Lengthened Interstimulus Interval: Kenneth R. Burstein*: Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Transfer of Training across Target Sizes: Ina McD. Bilodeau*: Department of Psychology, Tulane University, New Orleans 18, Louisiana.

Effect of UCS Strength on GSR Conditioning: A Within-Subject Design: Delos D. Wickens* and Gordon B. Harding: Department of Psychology, University of California, Berkeley, California 94720.

The Primacy of a Form Criterion in Perceptual Judgments: Albert S. Rodwan*: Department of Psychology, Emory University, Atlanta, Georgia 30322.

Short-Term Memory under Work Load Stress: Robert Seibel,* Richard D. Christ, and Warren H. Teichner: Institute of Environmental Psychophysiology, Clark Hall, University of Massachusetts, Amherst, Massachusetts 01003.

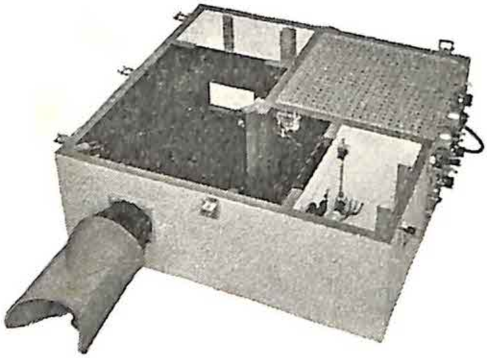
Graded Contrast Effects in the Judgment of Lifted Weights: Vincent Di Lollo* and J. H. Casseday: Department of Psychology, Indiana University, Bloomington, Indiana 47405.

Tachistoscopic Recognition Thresholds, Paired-Associate Learning, and Free Recall as a Function of Abstractness-Concreteness and Word Frequency: Wilma A. Winnick* and Kenneth Kressel: Department of Psychology, Queens College, Flushing 67, New York.

Function and Percentage of Occurrence of Response Members in Paired-Associate Learning: Robert K. Young* and Carl I. Fuhrmann: Department of Psychology, University of Texas, Austin 12, Texas.

Visuomotor Adaptation to Discordant Ex-afferent Stimulation: I. P. Howard,* B. Craske, and W. B. Templeton: Department of Psychology, University of Durham, Kepier Terrace, Gilesgate, Durham, England.

Human Probability Learning with Forced Training Trials and Certain and Uncertain Outcome Choice Trials: James K. Arima*: Army Combat Developments Command, Experimentation Center, Fort Ord, California 93941.



Journal of Experimental Psychology

Vol. 69, No. 2 February 1965

UNLEARNING AS A FUNCTION OF THE RELATIONSHIP BETWEEN SUCCESSIVE RESPONSE CLASSES[1]

LEO POSTMAN, GEOFFREY KEPPEL, and KAREN STARK

University of California, Berkeley

This study tests the assumption that in the A-B, A-C transfer paradigm the unreinforced evocation of covert or overt first-list responses during the acquisition of the second list results in the unlearning of the first-list associations. In an RI design using the A-B, A-C paradigm the work groups received interpolated trials on a second list in which the type of responses either remained the same as in List 1 or was different. Free recall of List 1 associations revealed greater RI and more interlist intrusions for the conditions using the same type of responses. It appears that the degree of unlearning depends upon the frequency with which first-list responses are elicited during second-list learning. These results are consistent with the position that unlearning is a process sharing some of the functional characteristics of experimental extinction.

Evidence for the unlearning of first-list associations (A-B) during the acquisition of a second list (A-C) has been presented by Barnes and Underwood (1959). An unpaced test of recall in which *S*s were required to report the responses from both lists showed that the number of first-list responses recalled declined steadily as a function of the degree of second-list learning. Since the conditions of recall minimized the opportunities for response competition, the progressive increases in retroactive inhibition (RI) were attributed to the gradual weakening of the first-list associations during interpolated learning. In a later experiment, McGovern (1964) demonstrated that A-C learning not only weakens the A-B or forward associations but also reduces the availability of the first-list responses as such. Theoretical interpretations of these results have identified the unlearning of the verbal associations with the process of extinction in classical conditioning. Thus, it is assumed that covert or overt first-list responses are evoked during the acquisition of the second list and that as a consequence of their failure to receive reinforcement the first-list associations undergo extinction.

[1] This research was carried out under Contract Nonr 222(90) between the University of California and the Office of Naval Research. This work was done at the Institute of Human Learning which is supported by a grant from the National Science Foundation.

Although the concept of unlearning represents an important step in the analysis of the mechanisms of transfer in verbal learning, there have been few investigations of the boundary conditions within which this analysis applies. Thus, studies investigating unlearning in the A-B, A-C paradigm have typically used nonsense-syllable stimuli and adjective responses in both lists. In these studies, and in fact in all experiments on unlearning thus far, the B and C responses learned in the successive lists are from the same form class, e.g., adjectives. Within the framework of a theoretical analysis which identifies unlearning with extinction it becomes an important question, however, whether the amount of unlearning varies with the similarity of the form class of the successive response terms.

As stated earlier, first-list responses are assumed to undergo extinction as a result of their unreinforced evocation as implicit or explicit errors during second-list learning (cf. Melton, 1961; Melton & Irwin, 1940). It is possible, however, that unlearning is independent of the evocation of the first-list response and reflects the displacement of an old response by a new one. These two interpretations lead to different predictions about the effects which manipulation of the form class of the responses should have on the amount of unlearning. It is reasonable to suppose that the probability of implicit and explicit interlist errors varies directly with the similarity of the form class of the successive response terms. If the degree of unlearning depends on the frequency with which first-list responses are elicited during the acquisition of the second list, the amount of RI should then be greater when the form class remains the same than when it changes. On the other hand, if unlearning depends merely on the acquisition of a new response to an old stimulus, the amount of RI should be independent of the similarity of form class.

It is interesting to note that the prediction based on a theory of extinction is exactly opposite to that implied by Osgood's (1949) analysis according to which RI increases as the similarity of the response terms decreases. Since responses from the same form class are more similar than responses from different form classes, greater RI should be found under the latter condition.

In the present experiment *S*s learned two lists in which the stimuli (numbers) remained identical. The response terms in the two lists either came from the same class (adjective-adjective or letter-letter) or from different classes (adjective-letter or letter-adjective). This arrangement permitted a test of the experimental hypotheses stated above, and at the same time provided an opportunity to determine the generality of the unlearning phenomenon with new types of materials.

METHOD

Design.—The design comprised two RI work conditions which differed with regard to the similarity of the form class of the two sets of response terms, and a control condition. The work groups learned two lists of paired associates (A-B, A-C). One work condition (Cond. S) received the same type of response terms in the two paired-associate lists and the other condition (Cond. D) received different types of response terms in the two lists. The control groups learned a single list. A second variable, type of first-list pairs (number-adjective or number-letter), was combined factorially with the three conditions. Thus, the complete experimental design consisted of six different treatment combinations.

Materials.—Each list consisted of eight paired associates in which the stimuli were the single-digit numbers 2–9 and the responses were either adjectives or letters. The

two lists which were presented to the RI groups conformed to the A-B, A-C transfer paradigm. The 16 response words were two-syllable adjectives, chosen for minimal similarity from Haagen's (1949) list of adjectives and divided into two sets of comparable Thorndike-Lorge (1944) frequency. The 16 letters were selected by eliminating from the alphabet the three most frequent letters, all vowels, letters corresponding to Roman numerals, and letters for which confusion with other letters may occur. The 16 letters were divided into two sets which are comparable in frequency of occurrence in printed text. The actual response units consisted of repeated letters, e.g., GG or ZZ. This was done to equate the time required for the pronunciation of the two-syllable adjectives and the letters. For both form classes, the two subsets of responses were used equally often as first-list responses.

In Cond. S in which the two sets of response terms were from the same form class, the similarity between the two responses which would be associated with a common stimulus was minimized for both the adjectives and letters. For the adjectives the pairing was random with the exception that similar suffixes, similar meanings, and initial letters that are adjacent in the alphabet were not allowed. The pairing of the letters was such that the paired letters did not rhyme, were not adjacent in the alphabet, and did not suggest abbreviations or words. For Cond. D two restrictions were placed upon the pairings of adjectives and letters: the letter could not occur in the paired word and the letter could not be adjacent in the alphabet to the initial letter of the paired word. For word-word, letter-letter, and word-letter pairs, two such pairings were employed. The particular stimulus-response pairings were chosen to avoid similarities in sound and appearance. Finally, four different orderings of the pairs were used to minimize serial learning of the responses and to provide four different starting positions.

For purposes of group designation two letters will describe the experimental treatment. The first letter of the designation (A or L) will refer to the type of response term, adjective or letter, during first-list learning. The second letter will be assigned on the same basis to indicate the type of response term during second-list learning, with the letter R (rest) used to designate the control condition. Thus, Groups A-A and L-L comprise Cond. S, and Groups A-L and L-A comprise Cond. D. Groups A-R and L-R are the control groups learning adjectives and letters, respectively.

Procedure.—Both lists were learned by the anticipation method at a 2:2-sec. rate; the intertrial interval was 4 sec. The first list was presented to a criterion of one perfect recitation. Approximately 1 min. following the attainment of the criterion all *S*s were given the instructions for the interpolated activity. For the RI groups this activity consisted of a second list of paired associates which was presented for 20 anticipation trials, while for the Control groups a series of arithmetic problems was administered. At the end of the interpolated activity (18 min. following the end of first-list learning) both groups were given a modified free-recall test (MFR). The MFR test consisted of an unpaced presentation of each of the eight stimulus terms; to the right of each stimulus item was a dash indicating the missing first-list response and the appropriate second-list response. It was made clear to *S* that the second-list response would be provided in each case. The *S* was given unlimited time to give the correct first-list response for each stimulus, the drum being advanced whenever a response was given or whenever *S* indicated an inability to recall. The order of the stimulus terms on the MFR test corresponded to the order of the pairs on the trial following the first-list criterion trial.

It should be noted that the present experiment represents a departure from the procedure employed by Barnes and Underwood (1959) to study the unlearning of first-list associations. These investigators required *S*s to write down the response terms from both paired-associate lists. At the higher degrees of second-list learning there was a tendency for the *S*s to give the second-list responses first. This order of recall has two possible consequences: it provides an additional unlearning trial and effectively lengthens the retention interval for the first-list responses. In the present experiment the effect of this "interpolated recall" was eliminated by limiting the test of retention to the first-list responses and by providing the correctly paired second-list responses at the time of recall.

Subjects.—A total of 96 *S*s, 16 in each of the six conditions, participated in the study. The *S*s were undergraduate psychology students at the University of California. The experimental conditions were randomized into 16 blocks of 6, each condition appearing once in each block. The particular list order, list pairing, and starting order were assigned

TABLE 1

Learning and Recall Measures for the Various Groups

	List 1 Form Class					
	Letters			Adjectives		
	Group L-L	Group L-A	Group L-R	Group A-A	Group A-L	Group A-R
List 1: Mean trials to criterion	13.8[a]	14.3	11.1	10.9	10.4	11.4
List 2: Mean numbers of correct responses	120.6[b]	131.4	—	139.1	126.8	—
MFR: Stringent	4.75[c]	7.06	7.69	4.75	6.38	7.56
MFR: Lenient	5.19[d]	7.44	7.81	4.81	6.44	7.56

[a] $MS_w = 34.17$.
[b] $MS_w = 405.23$.
[c] $MS_w = 1.80$.
[d] $MS_w = 1.43$.

randomly to the 16 blocks with the restriction that these assignments be balanced over the 96 *S*s. The *S*s were assigned to the randomized order of the conditions in order of their appearance in the laboratory. There were no restrictions placed on the selection of *S*s, most of whom had participated in previous verbal-learning experiments. No *S* was eliminated for failure to reach the first-list criterion.

Results

First-list learning.—The mean numbers of trials to a criterion of one perfect recitation on List 1 may be found in the first row of Table 1. In this table the six groups are represented by the column entries, the first three columns consisting of groups receiving letters as first-list response terms, the second three columns consisting of groups receiving adjectives as first-list response terms. Since the variances were homogeneous for all measures, the error mean square is given at the bottom of the table. The two work groups learning letter responses reached criterion somewhat more slowly than the remaining groups. Analysis of variance shows, however, that the variation in speed of List 1 learning is not significant ($F = 1.22$), nor is there a reliable overall difference between the two types of list, $F\ (1, 90) = 3.17, p > .05$. It should be noted that the average speed of learning is virtually identical for the two groups comprising Cond. S (A-A and L-L) and Cond. D (A-L and L-A), respectively.

Second-list learning.—The mean numbers of correct responses (out of a possible 160) during the 20 anticipation trials on List 2 are shown in the second row of Table 1. Adjective responses again have an advantage over letter responses, and this difference reaches statistical significance in List 2 learning, $F\ (1, 60) = 5.31, p < .025$. The degree of List 2 learning for Cond. S and D is almost exactly the same ($F < 1$). The interaction of type of list (A vs. L) with the condition of interpolation (S vs. D) is not significant ($F = 1.92$). It is to be noted, however, that when the List 2 responses are adjectives, Cond. S surpasses Cond. D (A-A > L-A) whereas the opposite is true when the List 2 responses are letters (A-L > L-L). As Fig. 1 shows, this interaction is clearly present on the early trials of List 2 learning, and it is statistically significant when the first two interpolated trials are considered, $F\ (1, 60) = 13.00, p < .01$. This find-

ing suggests that identity of response class may have two opposed effects on the speed of List 2 learning: (*a*) the set to give the appropriate class of responses facilitates performance; (*b*) failures to discriminate the list membership of specific items within the common response class is conducive to interlist interference. The more limited the pool of items in the common class the more likely it is that such interlist interferences will outweigh the facilitative effects of response restriction. The total pool of letters is, of course, considerably more limited than that of adjectives. Hence the net difference between Cond. S and D is positive for adjectives and negative for letters, especially on the early trials of List 2 learning when response restriction and list differentiation should have their greatest effects. This interpretation is supported by an examination of the interlist intrusions in List 2 learning.

Interlist intrusions during interpolated learning.—Neither of the groups in Cond. D gave any interlist intrusions during List 2 learning. There were two intrusions contributed by two *S*s in Group A-A and six intrusions contributed by four *S*s in Group L-L. Although the absolute frequencies of intrusions are as usual very small, the difference between the groups is in accord with the assumption that interlist differentiation increases with the size of the response pool. All but one of these intrusions occurred during the first six trials of interpolated learning.

Recall.—The last two rows of Table 1 show the mean numbers of correct List 1 responses on the MFR test. A "stringent" score and a "lenient" score are presented in each case. The stringent scores are based on responses given to the appropriate stimulus. The lenient scores are based on all

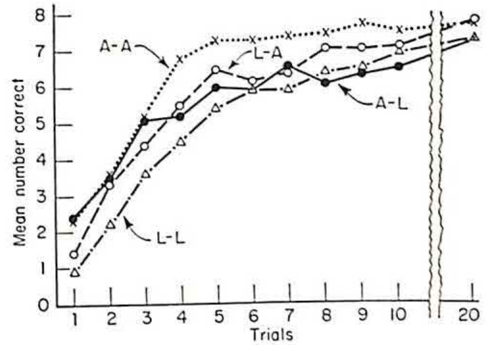


FIG. 1. Mean numbers of correct responses in interpolated learning. (There was little change in performance between Trials 11 and 19.)

List 1 responses recalled, regardless of whether or not they were given to the appropriate stimulus.

The stringent scores will be considered first. The two control groups, A-R and L-R, have very similar scores and show only a small amount of forgetting. Each of the four work groups recalls fewer List 1 responses than its corresponding control group. The overall difference between the control and work treatments is highly significant, $F(1, 90) = 42.37, p < .001$. Thus, there is clear evidence for unlearning of first-list associations. The amount of unlearning, as measured by the control-work differences, does not vary reliably with the type of response term ($F < 1$). For both types of list the retention losses are, however, substantially greater in Cond. S (A-A and L-L) than in Cond. D (A-L and L-A), $F(1, 90)$ for the difference between the two conditions of interpolation is 34.45 ($p < .001$). Again this difference does not vary with type of list ($F < 1$).

With one exception (A-R), the lenient scores of all groups are higher than the stringent scores. These increases reflect misplaced responses which were not also given to the appropriate stimulus terms. For each

TABLE 2

INTRALIST ERRORS AND INTERLIST INTRUSIONS IN MFR

	List 1 Form Class											
	Letters						Adjectives					
	Group L-L		Group L-A		Group L-R		Group A-A		Group A-L		Group A-R	
	f	n	f	n	f	n	f	n	f	n	f	n
Intralist	15	8	10	7	4	2	5	4	4	2	2	2
Interlist	13	8	0	0	—	—	2	2	0	0	—	—

Note.—f = total frequency of errors; n = number of *S*s contributing errors.

treatment, the group learning letter responses shows a greater gain than the group learning adjective responses. This differential gain results in a significant overall difference between the two types of lists in the amount recalled as measured by the lenient scores, $F(1, 90) = 4.92$, $p < .05$. In all other respects the pattern of differences, and notably the separation between Cond. S and D, are the same for the lenient and the stringent scores, and are associated with the same levels of statistical significance.

Errors in MFR.—Since the stimulus terms accompanied by the appropriate List 2 responses were exposed one by one in MFR, it was possible for *S* to make both intralist errors (misplaced responses from within List 1) and interlist errors (inappropriately paired List 2 responses). The total frequencies of such errors and the numbers of *S*s contributing each are shown in Table 2. It is clear that intralist errors occur more frequently in the recall of letters than of adjectives. This finding is brought out most clearly in the comparison between Groups A-A and L-L since the numbers of correct responses and hence the opportunities for errors were identical in these two cases. The difference between Groups A-L and L-A is in the same direction and remains intact when opportunities for errors are taken into account. Interlist errors at recall occur in Cond. S but not in Cond. D and again are substantially more numerous in the recall of letters than of adjectives.

DISCUSSION

The results of the experiment show significantly greater unlearning of the first list in an A-B, A-C paradigm when the class of responses in the successive lists remains the same than when it changes. The difference between Cond. S and D does not interact with the type of list, i.e., it is present to a comparable degree in the recall of both letters and adjectives. These findings are in accord with the hypothesis that the amount of unlearning depends on the frequency with which List 1 responses are elicited as covert or overt errors during List 2 learning. Thus, the interpretation of unlearning as a process sharing some of the functional characteristics of extinction appears to receive additional support.

The validity of this conclusion depends, of course, on the correctness of the assumption that the probability of unreinforced elicitations of List 1 responses depends on the similarity of form class. There is substantial evidence for the operation of a selector mechanism in learning and recall (Underwood & Schulz, 1960, pp. 143 ff.), i.e., for *S*'s tendency to restrict his responses to the categories

of items represented in a list. It is reasonable to suppose, therefore, that during List 1 learning *S*'s range of responses becomes restricted to the class of items appropriate to that list. If the form class remains the same in List 2, *S*'s responses will continue to be drawn from the same restricted range; to the extent that there are failures of differentiation among items within the appropriate class, List 1 responses will occur as implicit or explicit errors. A change in form class leads to a new basis of response selection during acquisition of List 2, and hence to a reduction in the probability of interlist errors. The difference between Cond. S and D in the frequency of overt intrusions during List 2 learning conforms to these expectations. The fact that there is some unlearning under Cond. D indicates, however, that there are limitations on the effectiveness of the selector mechanism in protecting List 1 responses from being elicited as errors and undergoing extinction. It is possible, of course, that extinction produced by unreinforced evocation of List 1 responses is not the only factor contributing to failures of List 1 recall. For example, performance in MFR may not be completely free from the effects of response competition, i.e., dominant List 2 responses may sometimes interfere with List 1 recall even without prior extinction.

Although the variations in the amount of unlearning are parallel for the two types of list, it is clear that interitem differentiation is more difficult for letters than for adjectives. Group L-L gave more interlist intrusions than Group A-A both during List 2 learning and in MFR. In addition, the numbers of misplaced List 1 responses were greater in the recall of letters than of adjectives. This difference has been attributed to the fact that the total pool of letters is considerably smaller than that of adjectives. Preexperimental associations are likely to be strong among the members of a strictly circumscribed class of responses. Thus, letters are used continuously in different combinations and there are strong sequential dependencies among them. As a result, the members of this class will tend to elicit each other in learning and recall. There are no comparable preexisting associations among randomly selected members of the broad class of adjectives. Moreover, the smaller the pool of items the more readily available all possible responses are to *S*. The tendency to guess may be expected to increase with the availability of the total range of responses. Guessing will, of course, increase the frequency of overt errors.

There is no systematic relationship between amount of unlearning and error rate in MFR. Although the numbers of misplaced responses and interlist errors are higher in the recall of letters than of adjectives, the stringent scores for Groups L-L and A-A are identical and Group L-A has an advantage over Group A-L. It appears that overt errors occur when the correct responses are not available, and that they do so at different rates depending on the class of responses. If this interpretation is correct, misplaced responses and interlist intrusions in MFR are a consequence rather than a determinant of failures to recall the responses to specific stimuli (cf. Conrad, 1960).

A high rate of intralist errors in MFR will also serve to enhance the differences between the stringent and lenient scores. Thus, while the stringent scores for the two types of list do not differ significantly, the lenient scores are reliably higher in the recall of letters than of adjectives. For purposes of interpreting this pattern of differences it may be assumed that (*a*) the lenient scores measure the strength of the association of the responses to the environmental context, whereas (*b*) the stringent scores reflect in addition to association with context the strength of specific S-R connections (cf. McGovern, 1964; Underwood, Keppel, & Schulz, 1962). It follows that the rate of unlearning of specific S-R associations is independent of the type of response whereas the association to contextual stimuli is more resistant to extinction for letters than for adjectives. This interpretation must

be viewed with caution, however, since the number of letter responses may well be inflated by guessing. In any event, the comparability of the stringent scores for the two classes of responses extends the generality of the unlearning phenomenon to include both letter and adjective responses to numbers.

The results of the experiment have general implications for the role of response similarity in transfer and interference. Reference has already been made to Osgood's (1949) analysis according to which negative transfer and RI are expected to increase as successive responses to the same stimuli become more dissimilar. This relationship does, indeed, hold true when degree of response similarity is varied within the same response class. Thus, with responses drawn from the same class, A-B, A-C (identical stimuli and unrelated responses) is a paradigm of negative transfer, and A-B, A-B′ (identical stimuli and similar responses) is a paradigm of positive transfer. The present findings indicate that this relationship does not apply to the similarity of response classes and instead suggest a new principle: the more dissimilar the classes of responses attached to the same stimuli the more likely it becomes that the successive response systems will maintain their strengths independently of each other.

REFERENCES

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Conrad, R. Serial order intrusions in immediate memory. *Brit. J. Psychol.*, 1960, **51**, 45–48.

Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

McGovern, J. B. Extinction of associations in four transfer paradigms. *Psychol. Monogr.*, 1964, **78** (16, Whole No. 593).

Melton, A. W. Comments on Professor Postman's paper. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 179–193.

Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

Osgood, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Underwood, B. J., Keppel, G., & Schulz, R. W. Studies of distributed practice: XXII. Some conditions which enhance retention. *J. exp. Psychol.*, 1962, **64**, 355–363.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received January 7, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 119-120

FUNCTION ORDER AND PAIRED-ASSOCIATE LEARNING [1]

CAMERON R. PETERSON, Z. J. ULEHLA, AND RICHARD S. LEHMAN

University of Colorado

A paired-associate learning task involved functions which related numerical responses to numerical stimuli. Some order characterized each of 17 different functions. The results showed that Ss used the order incorporated in a function to reduce their error in responding to stimuli encountered for the first time. For all different functions, error in initial responses to new stimuli decreased as the number of experienced S-R pairings increased.

In paired-associate tasks characterized by orderly relations between the stimulus and response variables, it is possible that *S*s can use the order in earlier experienced S-R pairings to aid them in responding to new stimuli. The purpose of the present experiment is to determine whether *S*s use such order to decrease error in responding to a stimulus which has not yet been experienced.

To illustrate, suppose that an *S* has experienced the following four S-R pairings: 1-11, 2-12, 3-13, and 4-14. The *S* would very likely make the appropriate response, 15, to the stimulus 5 even though he has never before experienced it. Although the order in this illustration is simple to the point of being trivial, the possibility exists that *S*s can use more subtle types of S-R order to reduce error in responding to new stimuli. The purpose of this experiment is to see if error in initial responses to new stimuli decreases as the number of experienced S-R pairings increases, when responses are related to stimuli by some orderly function. Such improvement would indicate a type of generalization or transfer of training which is mediated by abstracting the order in the function relating responses to stimuli.

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 40 of the Institute. This research was supported by a research grant (M-4977) from the National Institute of Mental Health. Grateful acknowledgment is made by the authors to K. R. Hammond and F. J. Todd for valuable suggestions.

METHOD

Subjects.—One hundred thirty-five students from an introductory psychology course at the University of Colorado served as *S*s.

Stimuli, responses, and functions.—The ordered set of integers 11–36 served as the stimulus variable and the set of integers 37–99 served as the response variable. Seventeen different functions paired a response with each of the stimuli. Nine of the functions were sampled from mathematics texts and included such functions as a straight line, a sine curve, an exponential curve, and a normal curve. Since only integers were used, discrete approximations to the theoretically continuous functions were employed.

The remaining functions were created to incorporate varying degrees of order. They are presented in Table 1. The left-hand column lists stimulus values; each of the other columns indicates the associated response value for one of the functions. All functions embody some order. The type of order becomes intuitively apparent when they are graphed.

Procedure.—Each *S* was assigned to a single function. The stimulus and response values were placed on a long strip of paper to be run through a typewriter. The *S* was presented with a stimulus number, typed his guess of the response number, and advanced the paper two spaces which allowed him to see the correct response number. Advancing

TABLE 1

S-R VALUES FOR EACH FUNCTION

Stimulus Values	Functions							
	1	2	3	4	5	6	7	8
11	42	42	42	42	91	42	44	42
12	43	43	43	43	95	46	46	45
13	45	44	44	47	99	50	48	43
14	46	47	47	45	98	47	50	47
15	48	50	50	46	97	44	47	50
16	49	51	51	50	91	46	44	48
17	51	52	54	48	85	48	41	52
18	52	55	55	49	89	52	45	50
19	54	58	56	53	93	56	49	54
20	55	59	59	51	92	53	53	57
21	57	60	62	52	91	50	55	55
22	58	63	65	56	85	52	52	59
23	60	66	66	54	79	54	56	62
24	61	67	67	55	83	58	58	65
25	63	68	68	59	87	62	60	69
26	64	71	71	57	86	59	62	67
27	66	74	74	58	85	56	59	71
28	67	75	75	62	79	58	56	69
29	69	76	78	60	73	60	53	72
30	70	79	79	61	77	64	57	70
31	72	82	80	65	81	68	61	74
32	73	83	83	63	80	65	65	72
33	75	84	86	64	79	62	67	75
34	76	87	89	68	73	64	64	79
35	78	90	90	66	67	66	68	82
36	79	91	91	67	71	70	70	80

the paper one more space brought the next number into view. The typewriter was equipped with a shield which prevented *S* from seeing more than one previous trial.

The stimulus sequence, identical for all functions and all *S*s, was generated by a random process. The *S* was required to work until (*a*) he completed all of the available 598 trials, or (*b*) the 2 hr. available for the experiment elapsed, or (*c*) he completed a series of 30 consecutive correct responses. Data analyses assume that those *S*s who made 30 consecutive correct responses would have responded correctly to all the remaining stimuli.

Data analysis.—Six to eight *S*s were employed for each function. Seven *S*s were discarded for failure to progress far enough through the trials to observe each of the 26 different S-R pairings at least once. This left 128 *S*s for the analysis.

If *S*s capitalize on S-R order, then error associated with responses to new stimuli will decrease as the number of experienced S-R pairings increases. Thus, if *S*s use order to extrapolate or generalize from experienced S-R pairings to novel pairings, the slope of the least-square straight line relating average error over *S*s for the first occurrence of each stimulus to the ordinal position of the first occurrence will be negative. This slope was obtained for each of the 17 functions.

RESULTS AND CONCLUSION

The mean slope of the least-square line relating average error for the first occurrence of each stimulus to the ordinal position of the first occurrence was −.308; the slope for each of the 17 functions was negative; and the 95% confidence interval for the mean slope was bounded by −.206 and −.410. Thus average response error to new stimuli decreased about .3 unit for each additional S-R pairing experienced. Despite their wide variety, some decrease characterized all functions. The conclusion is that *S*s do use order embodied in S-R relations to aid them in responding to new stimuli.

Additional analyses measured the relations between (*a*) degree of order characterizing the functions and (*b*) paired-associate performance on the functions. None of the measures of *a* predicted *b* very well.[2] Yet, since *S*s do use order in S-R relations, the quantification of such order remains an important, but unsolved, problem.

[2] These analyses are contained in Behavior Research Laboratory Report No. 39, Institute of Behavioral Science, University of Colorado.

(Received December 20, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 121–125

RATE OF PRESENTATION IN SERIAL LEARNING

GEOFFREY KEPPEL AND ROBERT J. REHULA

University of California, Berkeley *Science Research Associates, Chicago*

Serial-learning performance was studied under 2 rates of presentation, 2 and 4 sec. Following the attainment of criterion (5/14 and 10/14) presentation rate was either switched to the other rate or not switched. A final condition consisted of an alternation of 2- and 4-sec. trials. Although criterion was reached more rapidly under the 4-sec. rate, the total learning time required did not differ between the 2 rates. Performance on the postcriterion trials was primarily a function of the rate of presentation on the postcriterion trials with the precriterion rates having limited influence over the postcriterion trials.

Bugelski (1962) varied the study interval in a paired-associate task from 2 to 15 sec. and found a direct relationship between learning rate and length of the study interval. However, total learning time was constant (Rate × Trials-to-Criterion) for the various groups. Although there have been no studies reported for serial learning where rate of presentation has been varied widely, there is some indication that a similar constancy holds over a restricted range of rates (e.g., Hovland, 1938; Melton & Stone, 1942). Moreover, there appears to be a limit to this constancy, Braun and Heymann (1958) reporting the invariance for high-meaningful paralogs, but not for low paralogs.

The purpose of the present experiment was to determine whether the attainment of a common criterion under different rates of serial presentation reflects equal degrees of learning. In the traditional method of serial-anticipation learning, the study and testing intervals are unavoidably confounded, variation in presentation rate resulting in the variation of both of these intervals. Under these circumstances equal performance may not reflect equal learning if the responses recalled represent differences in latency. That is, if some of the items which were correct at the slower rate represent latencies which are longer than the length of the interval for the faster rate, these items may be said to reflect less strength than an item which is anticipated within the shorter interval. To the extent that performance at a slower rate consists of longer latency responses, attainment of the same criterion does not insure equal degrees of learning. If this is the case, conclusions concerning the constancy of total learning time under different rates must be rejected for serial learning.

In this study *S*s learning under two rates of presentation, 2 and 4 sec., were taken to the same criterion level. Following the attainment of criterion, presentation rate was either switched to the other rate or not switched. If performance on the criterion trial reflects equal degrees of learning for the two rates of presentation, the two switched groups should adjust immediately to the postcriterion rate. On the other hand, if performance on the criterion trial reflects differences in responses latencies for the two rates of presentation, the slow-fast group should be inferior to the fast nonswitched group and the fast-slow group should be superior to the slow nonswitched group on the postcriterion trials.

Method

Design.—Two variables, presentation rate (2 and 4 sec.) and postcriterion condition (switched and nonswitched), were combined factorially at two different preswitch criteria (5/14 and 10/14). Independent groups of *S*s were employed for all switched conditions, while the four nonswitched conditions were represented by two groups both taken to the higher criterion, one for each of the two rates. A final condition consisted of a group receiving alternated fast-slow trials throughout the course of learning.

Procedure.—Precriterion training was given under either a 2- or 4-sec. rate of presentation. In the designation of conditions, the letters F and S will refer to the 2- and 4-sec. rates, the first and second letters to the pre- and postcriterion rates, and the subscripts L and H to the low and high criteria, respectively. Following the attainment of the criterion, the two nonswitched groups (Groups F-F and S-S) continued training at the precriterion rates and the two switched groups (Groups F-S and S-F) continued training at the reversed, postcriterion rates. The alternation group (Group Alt.) received the study trial at the 2-sec. rate, followed by reversal in rate (4, 2, 4, etc.) at the end of each anticipation trial. Postcriterion training consisted of an additional 10 trials at the postcriterion rate. For the nonswitched groups and Group Alt., the 10 transfer trials were administered following the attainment of the higher criterion. Thus, it was possible to employ these as comparison groups for the switched conditions at both criteria.

Standard serial anticipation learning was employed. The *S*s were not notified of the rate change, but were told to expect either a slowing down or a speeding up of the turning drum and that the rate change would be made during the intertrial interval. In addition, *S*s were asked to give their responses as quickly as possible at all times since they would receive no warning of the change in the rate of presentation.

Materials.—A single 14-adjective serial list, having high intralist similarity, was selected from the lists reported by Underwood and Goad (1951). An asterisk served as the anticipatory cue and the intertrial interval was either 2 or 4 sec. Since interest was not centered upon a precise determination of the serial-position curve, a single ordering of the adjectives was utilized.

Subjects.—Students enrolled in the introductory psychology classes at Northwestern University, for which service in experiments is a course requirement, served as *S*s in the experiment. Most of the *S*s had participated in previous paired-associate and serial experiments. The experimental conditions were randomized in blocks of seven and *S*s were assigned these conditions in the order of their appearance in the laboratory. Each group contained 15 *S*s. No *S*s were eliminated for failure to reach criterion.

Results

Precriterion Performance

Overall comparisons.—Except for different presentation rates all groups were treated alike on the precriterion trials. An analysis of the Rate × Switching factorial at each criterion provides a test of the rate variable and comparability of groups. At the lower criterion the only significant source of variability was presentation rate, $F\,(1, 56) = 10.41$, $p < .01$, the slower rate resulting in faster learning ($\bar{X} = 3.17$) than the faster rate ($\bar{X} = 5.63$). The same results were found at the higher criterion, $F\,(1, 56) = 36.02$, $p < .01$, the mean trials to criterion for the F and S groups being 15.17 and 7.87, respectively. However, this latter analysis also revealed a significant interaction between the two variables, $F(1,56) = 4.69$, $p < .05$, in which Group F-F reached criterion significantly faster than Group F-S, $t\,(56) = 2.37$, $p < .05$. Since the conditions were randomly assigned to *S*s, the difference between the two groups must be attributed to sampling error. Comparison of the F groups at the lower criterion indicates that Group F-F is deviant, requiring 4.7 trials to reach criterion while Groups F-S_L and F-S_H required 6.6 and 7.2, respectively. The influence of Group F-F will be minimized in the comparisons which follow by combining the various F groups in precriterion comparisons. Moreover, since none of the correlations between the criterion measure and postcriterion measures

are significant, the precriterion difference appears to be relatively unimportant in interpreting postcriterion data.

Learning-time constancy.—Although performance under the two rates of presentation required different numbers of trials to reach criterion, an inspection of means reveals a constancy in total learning time. Since in terms of total presentation time two 2-sec. trials are equivalent to one 4-sec. trial, the scores for the fast rate were divided by two and compared with those for the slow rate. At both criteria the corrected criterion scores for the combined groups did not differ between the two rates (F = .23 and .06).

In these comparisons and the ones reported by others (e.g., Braun & Heymann, 1958; Bugelski, 1962), learning-time constancy has been demonstrated only with the trials-to-criterion measure. At the start of Trial 1, *S*s in the S groups had studied each item for 4 sec. on the preceding study trial, while *S*s in the F groups had studied each item for 2 sec. However, total study time is equivalent at the start of Trial 2 since the F groups received an additional 2-sec. study on Trial 1. Thus, performance on Trial 1 for the S groups and on Trial 2 for the F groups allows a comparison of the two groups with total presentation time held constant. Other points of comparison are possible between consecutive S trials and even-numbered F trials. For the present experiment performance was compared over the first four comparison trials. The total number of correct responses was slightly higher for the S groups ($\bar{X}$ = 18.31 and 17.97), but this difference was not significant (F = .004). No consistent differences between the two rates over the four trials were evident (F = .15). In short, when comparisons of performance under different rates of presentation are arranged so that total study time is equivalent, no differences are observed in the number of responses correctly anticipated on these trials as a function of presentation rate.

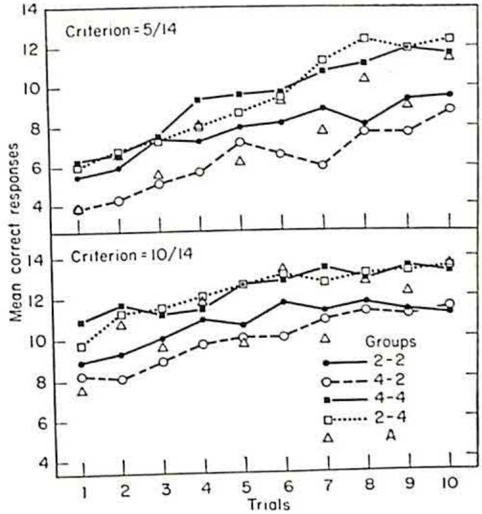


FIG. 1. The mean number of correct responses for the 10 transfer trials for the various groups following the attainment of the low and high criteria. (The group designations indicate the pre- and postcriterion presentation rates. Group A was administered the S rate on the even-numbered trials and the F rate on the odd-numbered trials.)

Postcriterion Performance

Postcriterion trials.—Following the attainment of either criterion, *S*s were given an additional 10 transfer trials at either the same or different presentation rates. The mean numbers of correct responses on the 10 transfer trials for the various groups are presented in Fig. 1. The triangles, unconnected for visual purposes, are for Group Alt. and will not enter into the main comparisons to be reported. Looking first at performance following the lower criterion, it appears that all groups except Group S-F are initially

equal in performance and that with additional trials Group F-F separates from the two S groups to merge with Group S-F. An analysis of the four groups in terms of the total number of correct responses summed across the 10 trials showed only the postcriterion rate to be significant, $F(1,56) = 30.76$, $p < .01$. An analysis over trials revealed a significant interaction with postcriterion rate, $F(9, 504) = 4.95$, $p < .01$, indicating a reliable difference in learning curves for the postcriterion rates. Other than Trials, none of the remaining comparisons was significant. Even though none of the comparisons including the precriterion rate was significant, Groups F-F and S-F were compared to determine the reliability of the differences reflected in Fig. 1. Neither a summation over trials ($\bar{X} = 77.60$ and 63.00) nor the interaction of groups and trials was significant, $F(1, 28) = .53$ and $F(9, 252) = 1.36$, respectively, $p > .05$.

The performance following the higher criterion is similar to that obtained following the lower criterion, with Group S-F again falling consistently below Group F-F and the S groups presenting indistinguishable learning curves. An analysis of the total number of correct responses revealed only the postcriterion rate to be significant, $F(1, 56) = 34.59$, $p < .01$; none of the interactions with trials was significant. A separate analysis of Groups F-F and S-F produced no significant differences for the comparisons of interest. In short, the analyses of the transfer trials after either criterion indicate that the only significant source of variability on the postcriterion trials is the rate which is being presented at that time.

Comparisons with Group Alt.—The performance of Group Alt. on the trials following the attainment of the two criteria is indicated in Fig. 1 by means of the unconnected triangles. Since in this condition presentation rate was shifted on each trial, some of the *S*s reached the criteria on the slow trial and some on the fast trial. Of the 15 *S*s in this group, 13 reached the criterion of 5/14 and 12 reached the criterion of 10/14 on the S trial. In order to give a clear picture of the performance of Group Alt. on the postcriterion trials, the data of the *S*s reaching criterion on the slow trial have been grouped and plotted in Fig. 1. Thus, on the odd-numbered transfer trials the list was presented at the F rate and on the even-numbered trials the drum turned at the S rate. The most salient aspect of the performance of Group Alt. is the fact that performance appears to be entirely determined by the rate of presentation on any given trial. To test this observation a comparison was made between Group Alt. and the combined switched and nonswitched groups on the trials for which the rate of presentation was the same. The results of these analyses were uniformly negative, the largest $F(1, 40)$ for the total number of correct responses summed over the five comparison trials being 1.13, $p > .05$, and the largest $F(4, 156)$ for the comparison of the interaction of the combined groups with Group Alt. over the five trials being 2.15, $p > .05$. It appears, then, that the performance of Group Alt. is entirely determined by the presentation rate given on any trial.

Discussion

The results of the present experiment corroborate the findings of others that rate of presentation in serial learning is an effective variable when measured in terms of trials to reach criterion, but does not produce differences in learning when measured in terms of total learning

time. Wider variations in rate and in the sampling of learning materials, however, will be necessary to determine the limits of this learning-time constancy for serial learning. In this regard, it is interesting to note that although presentation rate was not varied, a similar constancy was obtained by Waugh (1962) for different lengths of serial lists.

It was argued earlier that if the attainment of a common criterion implies equal degrees of learning under the two rates, a relatively fast and complete switchover to the postcriterion rate should occur. This is what appears to have happened: postcriterion performance was a function of the rate which was operating at that time. This conclusion obtains support from the postcriterion performance of Group Alt. where, after both criteria and for both rates of presentation, performance was perfectly appropriate to the current presentation rate.

REFERENCES

Braun, H. W., & Heymann, S. P. Meaningfulness of material, distribution of practice, and serial-position curves. *J. exp. Psychol.*, 1958, **56**, 146–150.

Bugelski, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 409–412.

Hovland, C. I. Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.

Melton, A. W., & Stone, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.

Underwood, B. J., & Goad, D. Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, 1951, **42**, 125–134.

Waugh, N. C. Length of series and the learning curve. *Amer. J. Psychol.*, 1962, **75**, 177–192.

(Received December 23, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 126–134

LEADERSHIP IN SMALL GROUPS: A MATHEMATICAL APPROACH [1]

ARNOLD BINDER
Indiana University

BURTON R. WOLIN AND STANLEY J. TEREBINSKI
System Development Corporation

The broad purpose of this research was the investigation of the relationship between decision-making success and the likelihood of being voted "leader" (group decision maker) of a 3-man group. Markov models, based on extensions of concepts used in mathematical learning theory, provided the theoretical framework. Each trial of the experiment began with the selection of a leader by group vote, and ended after the designated leader made a decision for the group. Obtained and predicted results were compared for voting shifts, asymptotic leadership and state proportions, and learning trends. 5 different reinforcement groups were run and in only 1 of these groups were there major discrepancies between actual and expected results.

The present research falls in the tradition set by Atkinson and Suppes (1958), Suppes and Atkinson (1960), Burke (1959, 1960), Estes (1962), and Hall (1962) in that it involves the analysis of small group behavior in terms of extensions of concepts and procedures from statistical learning theory.

Each group in this experiment contained three *S*s, and, on each of a series of trials, the group made a decision by the scheme which Smoke and Zajonc (1962) refer to as "dictatorship." But the individual who determined the response for the group on a given trial ("dictator") was chosen for that role by a majority vote of all members of the group. We shall refer to the individual who makes the decision on a given trial as the "leader," although he can exercise only a few of the attributes commonly associated with the leadership role. The broad purpose of the research is the investigation of the relationship between decision-making success and the likelihood of leadership selection in subsequent trials, using a mathematical model as the basis for analysis.

The model under consideration is Markovian. Let X, Y, Z designate each of the three group members, let $\pi_{i,n}$ be the probability that the decision of player i will be reinforced on Trial n if he is leader, and let $X_iY_jZ_k$ represent the states of the Markov process, where $i = y, z$; $j = x, z$; $k = x, y$. For example, the state $X_yY_zZ_y$ represents the condition where X designated Y as his leadership choice, Y designated Z, and Z designated Y. In this particular case, Y would become the elected leader. A random selection of leader was necessary in such states as $X_yY_zZ_x$ (which would occur by chance in $\frac{1}{4}$ of the trials).

We assume that the process of changes over trials in leadership voting involves probabilistic learning with reinforcement depending upon the correspondence between the leader's decision and that designated as

[1] This research was supported by the System Development Corporation where the senior author is a consultant to the Decision Processes Staff of the Research Directorate.

correct by the *E*. For this learning we assume a one-element pattern model with an axiom structure as summarized in Table 1 under the column "8-State."

Letting $\pi_{i,n} = \pi_i$ for all n, the transition matrix for this eight-state Markov process is readily obtainable: For example, the probability of the transition from $X_y Y_x Z_x$ to $X_y Y_x Z_x$ is $\pi_x + (1-c)^2(1-\pi_x)$ while the probability of the transition from $X_z Y_x Z_y$ to $X_y Y_x Z_y$ is $\frac{1}{3}c[\pi_y + (1-\pi_z)]$.

Since an eight-state Markov chain is cumbersome, it is convenient to introduce certain simplifying assumptions in order to reduce the number of states. The alternate axioms are summarized in the column "3-State" of Table 1. The transition matrix for the three-state process is shown in Table 2.

TABLE 1

ALTERNATIVE AXIOM SYSTEMS

Contingency	Probability of Shift of Member's Pattern	
	8-State	3-State
I The man designated by a particular member becomes leader. His decision is		
A Reinforced	0	0
B Not reinforced	c	c
II The man designated by a particular member does not become leader but the member himself becomes leader. The leader's decision is		
A Reinforced	0	.5
B Not reinforced	0	.5
III The man designated by a particular member does not become leader but the other man he could have designated does. The decision is		
A Reinforced	c	1
B Not reinforced	0	$(1-c)$

METHOD

Subjects.—A total of 102 male and female college undergraduates was used. These were recruited from various schools in the Los Angeles area and paid at an hourly rate.

Apparatus.—Each *S* sat at a table enclosed on three sides by heavy curtains which served to prevent *S*s from seeing the stimuli and responses of others. On the table before each *S* was an $8\frac{1}{2} \times 11$ in. cathode-ray tube display and a console with response buttons. Characters on the display tube were generated from a dot matrix. There were 24 such units in the experimental room with the necessary auxiliary equipment that made it possible to run eight groups (three *S*s to a group) simultaneously. Separate sets of response buttons were used for the leadership vote and for the leader's decision. Since *S*s were not permitted to vote for themselves, the button that would indicate a self-vote was covered. Indicator lights on the consoles were used to show voting and decision phases.

A Philco S-2000 computer, located in another room, generated the displays, timed the trials, recorded responses, and produced data print-outs following the experiment.

TABLE 2

THREE-STATE TRANSITION MATRIX

	X	Y	Z
X	$\pi_x + (1-\pi_x)\left(1 - \frac{5c}{3} + \frac{2c^2}{3}\right)$	$(1-\pi_x)\left(\frac{5c}{6} - \frac{c^2}{3}\right)$	$(1-\pi_x)\left(\frac{5c}{6} - \frac{c^2}{3}\right)$
Y	$(1-\pi_y)\left(\frac{5c}{6} - \frac{c^2}{3}\right)$	$\pi_y + (1-\pi_y)\left(1 - \frac{5c}{3} + \frac{2c^2}{3}\right)$	$(1-\pi_y)\left(\frac{5c}{6} - \frac{c^2}{3}\right)$
Z	$(1-\pi_z)\left(\frac{5c}{6} - \frac{c^2}{3}\right)$	$(1-\pi_z)\left(\frac{5c}{6} - \frac{c^2}{3}\right)$	$\pi_z + (1-\pi_z)\left(1 - \frac{5c}{3} + \frac{2c^2}{3}\right)$

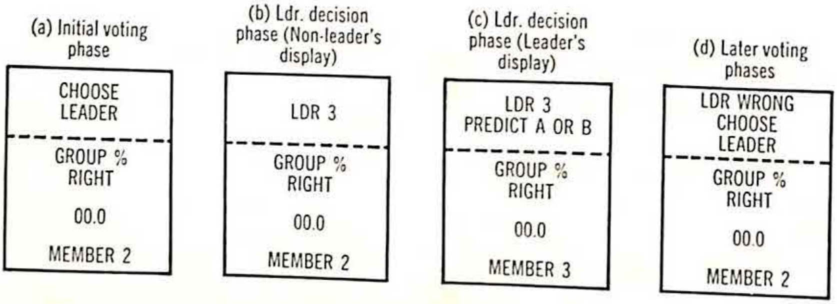


FIG. 1. Appearance of displays.

Three sets of punched cards, which were fed into the computer, controlled the reinforcement proportions of each group member when he acted as leader. Each set of punched cards consisted of a random sequence of reinforcements and nonreinforcements.

Procedure.—At the start of the experiment the information contained in Fig. 1a was on the scope for *S*s (note that the illustration is for the member whose number was 2; the member's number always appeared at the bottom of the display). This introduced the first or voting phase of Trial 1 which was defined as the period during which *S*s could vote for a leader. The *S*s were instructed to vote for one of the other two members of their group for leader who would in turn make a decision for the group as a whole. They were motivated by statements to the effect that the goal of each group was to make a high score (percent "Right" decisions) and that only the leader, as the group representative on any trial, had the opportunity to improve the score. A failure to vote on a given trial made the member ineligible to be selected as the leader on that trial.

The *S*s were assigned to the booths randomly and care was taken to prevent them from determining the other members of their groups during the experiment.

During the first 25 trials *S*s were given 6 sec. to vote for a leader. Beginning with the twenty-sixth trial the entire time cycle was speeded up and the time allowed for voting was reduced to $4\frac{1}{2}$ sec.

Following this voting phase came the decision phase of the trial, during which only the designated leader responded. Two responses were possible for the leader during this phase, predicting that either an A or a B was the correct decision by pushing the appropriate decision button on the console. The display which appeared during this phase on the scope of the two members who were not leader on the trial may be seen in Fig. 1b. The display of the leader for this phase differed only in that it contained instructions to predict A or B, in addition to the other information (see Fig. 1c). The leader was allowed 4 sec. to push the A or B button during the first 25 trials and 3 sec. thereafter.

The trial ended when the displays of the decision phase were removed from the scopes. The voting phase of the next trial was signaled by the appearance on all scopes of the information shown in Fig. 1d. The datum added on subsequent trials to the data on the scope during the voting phase of Trial 1 was the correctness ("Ldr Right") or incorrectness ("Ldr Wrong") of the leader on the immediately preceding trial. Moreover, the score of the group was altered (out to one decimal place) to confirm with the new datum. After the completion of this phase and a short delay came the leader-decision phase again with the displays as shown in Fig. 1b and 1c. All subsequent trials followed the same routine.

There was a delay of about $2\frac{1}{4}$ sec. between the end of a given phase and the disappearance of the scope information appropriate for that phase. This delay was necessitated by computer and programing characteristics.

The *S*s were run for 300 trials, which took about an hour, and then given a 10-min. rest. During the rest period *S*s were cautioned not to discuss the experiment and the display tubes were turned off so they could not determine group composition by group performance scores. Another 300 trials followed the rest period.

Random selection of a leader was made by the computer using a random number generator whenever: (*a*) all members voted and the result was a tie, (*b*) one member did not vote and the other two voted for each other or both voted for the ineligible member, or (*c*) none of the members voted.

The groups were run under five different reinforcement conditions, where a condition

TABLE 3

EXPERIMENTAL CONDITIONS

Reinforcement Cond.	Number of Groups
975	8
951	5
753	10
777	6
333	5

is defined by the set of π_i's for the given group. Table 3 shows the sets of reinforcement probabilities (conditions) used and the number of groups run under each. Reinforcement Cond. 975, for example, means that one member of the relevant group was reinforced (display indicating "Ldr Right") with a probability (p) of .9 for his decisions when leader, another member was reinforced with $p = .7$, and the third member with $p = .5$. For Reinforcement Cond. 777 and 333 the predictions of all three members were reinforced with the same probability, .7 or .3.

Reinforcement schedules were randomly determined in advance under the constraint that the given proportion hold exactly in groups of 20 trials. When a leader was scheduled for reinforcement (or nonreinforcement) it made no difference whether his response was actually A or B. A difficulty only arose when a leader failed to decide on a given trial, so that he was considered "Wrong" no matter what his reinforcement schedule called for. However, compensation was accomplished by designating as "Right" the very next decision he made as leader which was scheduled to be designated as "Wrong." Thus, it was possible to maintain a constant reinforcement proportion except toward the end of the experiment when an automatic "Wrong" might not be followed by a scheduled "Wrong."

Within all reinforcement conditions there was a counterbalancing of member's number (1, 2, or 3) with reinforcement probability to compensate for any response bias. The unequal numbers over the various conditions stemmed from equipment breakdowns as well as problems of S availability.

RESULTS

The relative frequencies of shifts in pattern conditions on given trials as functions of the status of the leader and the correctness of his decision on immediately preceding trials are shown in Table 4. The frequencies were computed over all Ss in each of the reinforcement groups. The data for cases in which S failed to vote or voted for himself on one of the relevant trials were ignored in the analysis for that particular trial. Erroneous voting occurred about 5% of the time.

The parameter can be estimated from I B or III A (Tables 1 and 4) for the eight-state process and from I B or III B for the three-state process.

TABLE 4

CONTINGENT VOTING BEHAVIOR OVER SS WITHIN REINFORCEMENT GROUPS

Contingency	Proportion Shifting Vote from Trial n to Trial $(n+1)$					
	777 Groups	333 Groups	951 Groups	753 Groups	975 Groups	Average All Groups
I						
A	.283	.232	.083	.097	.113	.162
B	.538	.390	.365	.325	.406	.405
II						
A	.291	.256	.113	.184	.119	.192
B	.301	.271	.246	.234	.192	.249
III						
A	.443	.431	.500	.499	.319	.439
B	.308	.254	.171	.197	.172	.220

Because of the desirability of using the same value in both matrices, I B was chosen for this purpose.

The formula for the stationary absolute probabilities of the three-state matrix is as follows:

$$u_i = \frac{(1-\pi_j)(1-\pi_k)}{(1-\pi_i)(1-\pi_j)+(1-\pi_i)(1-\pi_k)+(1-\pi_j)(1-\pi_k)}$$

where $i, j, k = X, Y, Z$ but $i \neq j \neq k \neq i$. However, it was not possible to obtain a general formula for the stationary probabilities of the eight-state matrix. Consequently, these probabilities were obtained by the process of raising the transition matrix to a high enough power for the necessary stability. The asymptotic proportions of leadership designation as a function of the probability of reinforcement for each member were derived from the last 200 trials of the experiment. These obtained values, together with the values expected on the basis of the stationary absolute probabilities of the three- and eight-state processes, may be seen in Table 5. The obtained values were calculated on two bases: first, using all trials no matter how the leader was selected, and second, excluding trials on which one or more members either did not vote or voted for themselves. For comparison purposes, the proportions expected on the basis of matching are also shown in Table 5; matching implies leadership selection at a relative frequency proportional to the probability of reinforcement.

The obtained asymptotic proportions (last 200 trials) for the states of the eight-state process, together with the predicted values, are shown in Table A.[2]

[2] Tables A and B and Figures A, B, and C have been deposited with the American Documentation Institute. Order Document

TABLE 5

Asymptotic Proportions of Leadership Status

Groups	Reinforcement	Obtained		Predicted		
		All Trials	With Exclusion	3-State	8-State	Matching
975	.9 .7 .5	.65 .23 .12	.67 .22 .11	.65 .22 .13	.65 .22 .13	.43 .33 .24
951	.9 .5 .1	.73 .21 .06	.75 .19 .06	.75 .15 .10	.76 .15 .09	.60 .33 .07
753	.7 .5 .3	.66 .26 .08	.71 .22 .07	.49 .30 .21	.49 .30 .21	.47 .33 .20
777	.7 .7 .7	.33 .32 .35	.34 .31 .35	.33 .33 .33	.33 .33 .33	.33 .33 .33
333	.3 .3 .3	.28 .36 .36	.28 .36 .36	.33 .33 .33	.33 .33 .33	.33 .33 .33

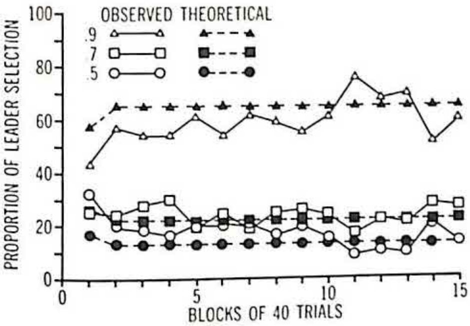


FIG. 2. Trends of leadership selection in 975 groups.

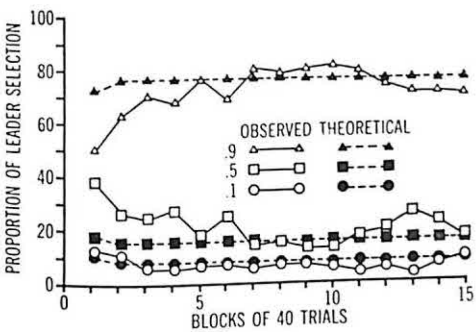


FIG. 3. Trends of leadership selection in 951 groups.

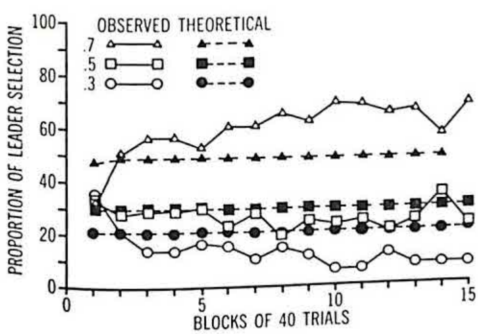


FIG. 4. Trends of leadership selection in 753 groups.

Now we turn to the learning process. Figures 2 to 6 show the trends of leadership selection in blocks of 40

No. 8133 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

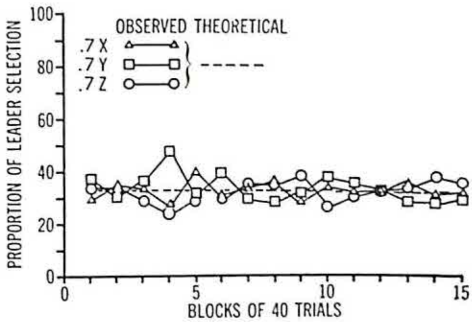


FIG. 5. Trends of leadership selection in 777 groups.

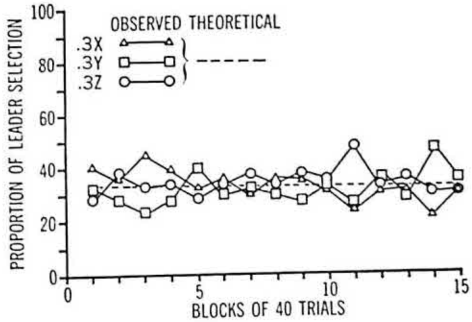


FIG. 6. Trends of leadership selection in 333 groups.

taken over all groups given a specified set of π's. These curves are based on the proportions each member was leader over the trials. The theoretical learning curves, based on the eight-state matrix, are also present in the figures.

DISCUSSION

While there would seem to be no reasonable basis for choosing between the eight-state and three-state matrices on the basis of asymptotic relative frequencies (Table 5), Table 4 does present relevant distinguishing data. Thus, using Table 1, in addition to Table 4, we note the following differences between the processes: the eight-state assumes the equality of the probabilities of shifts of patterns under Contingencies I B and III A, while the three-state assumes equality for the probability of shift under I B and the probability of remaining conditioned under III B; the

eight-state assumes no shift under Contingencies II A and II B, while the three-state assumes a probability of .5 for shift under both these contingencies; the probability of shift under Contingency III A is c for the eight-state and 1 for the three-state; and finally, the probability of shift under Contingency III B is 0 for the eight-state and $(1-c)$ for the three-state. In each of these comparisons the axioms of the eight-state process are more closely in accord with the data (using the averages in Table 4) than those of the three-state process, and in most cases considerably so. The relative frequencies of shift under Contingencies III A and III B (.439 and .220, respectively) are particularly telling. According to the eight-state process, the former should equal c (= .405 for I B) and the latter 0, while for the three-state, the corresponding values are 1 and $(1-c)$ which is equal to .595.

It may understandably be argued that even the eight-state model does not provide a very impressive fit in those cases where the theoretical probability is zero. One can only reply that the balancing of positive and negative discrepancies from the predicted value is not possible when the theoretical probability is zero and that nothing more than a first (and reasonable) approximation is claimed for the model. Moreover, the magnitude of the discrepancies from the $p = 0$ values is within the range found by other investigators in similar research areas; see for example Suppes and Atkinson (1960).

While the fit between obtained and predicted proportions of leadership designation (Table 5) as well as the overall theoretical and actual learning curves (Fig. 2–6) are generally close, a conspicuous set of failures occurs with Groups 753 in both asymptotic results and learning trends. A few other characteristics of Table 5 should be mentioned before these discrepancies are pursued further: First, the two processes predict almost identical asymptotic proportions, and second, the use of the data from all trials provides an excellent approximation to the more accurate (in terms of model axioms) analysis exclusively in terms of trials on which all *S*s voted.

The discrepancies for Groups 753 in Table 5 (as well as in Table A) clearly involve the choice of the $\pi = .7$ member as leader (from States $7_5 5_7 3_7$ and $7_3 5_7 3_7$) at the expense of the other states. Table B (see Footnote 2) contains an analysis of leadership choice among the 753 groups separately in an attempt to determine the generality of the disturbance. Seven of the 10 groups show marked overshooting in the choice of the .7 member as leader, and only two groups show proportions within reasonable distances of the prediction (.49). And in the two cases of adequacy for $\pi = .7$, the predictions for $\pi = .5$ and $\pi = .3$ are off by considerable margins. The data surely indicate that the miss between obtained and predicted (from the model) results in the 753 case is not an idiosyncrasy of one or a few groups which, if disregarded, would lead to adequate fit for the averages of the remaining groups.

In the comparison of predicted and obtained asymptotic proportions for the states of the eight-state process (Table A), it is interesting that the major source of discrepancy for the two other sets of groups where members did not have equal reinforcement probabilities (that is, Groups 975 and 951) is the tendency for the member getting reinforced at the highest rate (in these cases $\pi = .9$) to vote more frequently than expected for that member of the two remaining who has the higher reinforcement rate: that is, in the case of Groups 975, 9 votes too frequently for 7 rather than for 5; and in the case of Groups 951, 9 votes for 5 rather than for 1 more than expected. The high preference of 9 for choosing the individual most likely to get reinforced may stem from an unusually strong motivation for group success. It should be noted that this preference does not affect the relative frequencies of the alternate conditions of leadership, so that we find the close fit between predicted and obtained proportions (Table 5) for these groups.

Another way of evaluating the data is to consider that each S was faced with a discrimination problem which was more or less difficult depending upon the reinforcement ratios of his group as a whole and upon his specific reinforcement ratio within the group. The differential reinforcement any member received for voting for one member as against the other, and whether his own decision reinforcement was above or below the chance level of .5 might well have influenced voting behavior.

Members of the 975 and 753 groups had a more difficult discrimination to make than members of the 951 groups. Within those three groups, the members with the medium reinforcement ratio should have less difficult discriminations than the members receiving the high or low reinforcements. However, when the differences in reinforcement ratios are relatively large, as in the 951 groups, all members have a relatively easy discrimination task.

By recasting the data in Table A to show differential voting behavior of members of the various groups rather than occurrence of states it is possible to examine the effects of these factors. This has been done in Table 6.

This table was constructed by separately summing up the proportions for the four states by reinforcement member where he voted for the more successful member and for the four states where he voted for the less successful. Assuming that the member who is reinforced most frequently for his decisions has the greatest motivation for group success, we would expect overshooting toward higher reinforcement members to show up most clearly in the case of Groups 951. This is confirmed in Table 6 by the large overweighting of the preference of the high-reinforcement member in favor of the medium- (rather than the low-) reinforcement member for 951 as compared with 975 and 753. That a similar tendency may not be present for medium- and low-reinforcement members is indicated by the relatively close fit between their obtained and predicted proportions in Groups 975 and 951.

TABLE 6

ASYMPTOTIC VOTING PROPORTIONS

Groups	Member	Other Members	Obtained	Predicted
975	.9	.7 .5	.746 .254	.557 .443
	.7	.9 .5	.785 .215	.769 .231
	.5	.9 .7	.775 .225	.720 .280
951	.9	.5 .1	.909 .091	.565 .435
	.5	.9 .1	.853 .147	.851 .149
	.1	.9 .5	.791 .209	.809 .191
753	.7	.5 .3	.587 .413	.554 .446
	.5	.7 .3	.865 .135	.654 .346
	.3	.7 .5	.736 .264	.603 .397

Again, 753 turns out to be pathological. The high-reinforcement member provides a reasonably good fit between obtained and expected proportions while the fit is very poor for the other two members. It is not readily apparent how the discriminations for the members of 753 are more difficult than those of 975.

One other aspect of the differential discriminations among reinforcement groups is indicated by the curves of Fig. A, B, and C (see Footnote 2) which show the voting preferences of members in blocks of 40 trials throughout the 600 trials. The members of 951 seem to arrive at their asymptotic values earlier than the others.

It is obviously the case that the model predicts a faster rate of acquisition than is actually achieved. Considering the simplicity involved in the basic one-element pattern model, this result is not surprising. The rate at which a transition matrix reaches a stable state is a function of the accessibility of the states to each other and the variability of the

transition probabilities as well as the initial probability vector (see Doob, 1953). The theoretical learning process could be slowed down in the present case by shifting from the one-element pattern model to a multielement model and, also, by allowing for guessing states and states leading to no response. Future work will involve such effort.

It should be noted that only the group performance success score (percent "Right") was displayed. The *S*s had to calculate and remember the success ratios for the various group members without assistance. A separate score for each player might lead to even better correspondence of voting behavior to the model.

One modification of the experimental procedure which would lend itself readily to pattern considerations and might well modify the rate of approach to asymptote is to have *S*s vote for leader once every N trials, where N is large enough so that estimates of success rates of each member when leader could be made rapidly and accurately.

REFERENCES

Atkinson, R. C., & Suppes, P. An analysis of two-person game situations in terms of statistical learning theory. *J. exp. Psychol.*, 1958, **55**, 369–378.

Burke, C. J. Applications of a linear model to two-person interactions. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, California: Stanford Univer. Press, 1959. Pp. 180–203.

Burke, C. J. Some two-person interactions. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences: 1959.* Stanford, California: Stanford Univer. Press, 1960. Pp. 242–253.

Doob, J. L. *Stochastic processes.* New York: Wiley, 1953.

Estes, W. K. Theoretical treatments of differential reward in multiple-choice learning and two-person interactions. In J. H. Criswell, H. Solomon, & P. Suppes (Eds.), *Mathematical methods in small group processes.* Stanford, California: Stanford Univer. Press, 1962. Pp. 133–149.

Hall, R. L. Two-alternative learning in interdependent dyads. In J. H. Criswell, H. Solomon, & P. Suppes (Eds.), *Mathematical methods in small group processes.* Stanford, California: Stanford Univer. Press, 1962. Pp. 180–192.

Smoke, W. H., & Zajonc, R. B. On the reliability of group judgments and decisions. In J. H. Criswell, H. Solomon, & P. Suppes (Eds.), *Mathematical methods in small group processes.* Stanford, California: Stanford Univer. Press, 1962. Pp. 322–333.

Suppes, P., & Atkinson, R. C. *Markov learning models for multiperson interactions.* Stanford, California: Stanford Univer. Press, 1960.

(Received December 26, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 135–143

EFFECTS OF DIFFERENTIAL VALUE ON RECALL OF VISUAL SYMBOLS [1]

HARVEY A. TAUB [2]

University of Massachusetts

The effects of differential ratio of value, exposure time, and number of categories upon the detection and report of letters were investigated in a visual search task. The differential ratios of value of the letters were 2:1, 4:1, 8:1, and 16:1; the exposure times of the stimulus slides were 0.5, 1.5, and 2.5 sec.; and the numbers of categories on a slide were 4, 6, 8, 10, and 12 letters. The results indicated that Ss made more correct identifications, initial responses, and false reports of higher value symbols than of lesser value. Further, the percentage of correct identifications and percentage of initial responses of higher value symbols varied with exposure time and ratio. These effects were attributed to selective recall from short-term memory storage.

A study by Teichner, Reilly, and Sadler (1961) attempted to distinguish between perception and memory using a visual search task. The stimuli were slide projections which were exposed for a short period of time. Each slide contained a number (load) of different letters (categories) and repeated each category a number of times (density). The task of each *S* was to search, detect, and report the different categories which were displayed. It was assumed that perception included the search, detection, and storage of information (memory storage) which occurred during the exposure of the slide. Conversely, the concept of short-term memory was assumed to include the retrieval and report of letters from storage following the cessation of the slide. It was assumed further that the relative differences in number of letters perceived and the number reported may be investigated by varying the amount of information to be remembered (memory load), i.e., the less the memory requirement, the closer the approximation to pure perception or detection. The results of this study suggested that humans can receive more information than they can recall and that the percentage of correct identifications was inversely proportional to the number of categories (load). In general, these results agreed with the findings obtained with tasks which did not involve search (Anderson & Fitts, 1958; Sperling, 1960; Woodworth & Schlosberg, 1954). Recently, Teichner and Sadler (1962) further found that accuracy in the search task was a negatively accelerated, increasing function of exposure time with an asymptote at about 2.5 sec.

The present experiment was designed to extend the work of Teichner and Sadler (1962) by investigating the

[1] This paper is part of a thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Psychology, University of Massachusetts. The research was supported by funds from the Electronics Systems Division, Air Force Systems Command, Bedford, Massachusetts, under Contract No. AF 19(628)-290. A complete version of this paper was published as Report No. ESD-TDR-63-343. The author wishes to express his thanks to Warren H. Teichner, who was the principal investigator for the contract and who served as chairman of the author's advisory committee.

[2] Now at Cornell Aeronautical Laboratory, Inc.

effects of differential value of categories upon S's recall. More specifically, the experiment was concerned with the relative differences in accuracy of performance as differential value, memory load, and exposure time were varied. Investigations of decision-making situations (Tanner & Swets, 1954; Taub & Myers, 1961) suggested that in order to maximize performance, S's responses should vary directly with the relative value of the categories.

Method

Apparatus.—The stimuli consisted of 50 slides with white letters on a black background. The slides varied in number of categories (load) with each letter of the alphabet representing a category. The five levels of load were 4, 6, 8, 10, and 12 letters per slide. Each level of load was replicated 10 times with a different random sample of the alphabet. Density, the number of times a letter was repeated on each slide, was always constant at 2. To allow for a comparison of values, a distinction was made between the first and second halves of the alphabet. Each letter from A through K was worth one unit of value, while the letters P through Z had a different value. In order to provide a break between the two halves of the alphabet, the letters L, M, N, and O were not used on any of the slides. This provided two populations of category value, each containing 11 letters. All slides were designed to contain an equal number of letters from both halves of the alphabet. A description of the method of construction of the slides and a sample slide have been presented elsewhere (Taub & Teichner, 1963).

A Revere automatic slide projector with an externally mounted shutter presented the slides and two electronic timers controlled the exposure times and interslide intervals. The projected size of the slide was 40 × 60 in., while the projected size of the letters was 1 in. The Ss sat in student arm-type chairs, in groups of one to four, such that the average distance from their eyes to the projected letters was 96 in. The Ss performed under red illumination provided by two 25-w. light bulbs and reported their answers on prepared data sheets.

Procedure.—The 120 Ss were randomly assigned to 15 equal groups. The design is a 3 × 4 × 5 × 2 factorial with Exposure Time and Ratio of Value as "between effects" and Load and Category Value as "within effects." The exposure times were 0.5, 1.5, and 2.5 sec.; the four differential ratios of value were 2:1, 4:1, 8:1, and 16:1. Three equal-value (1:1) groups were run to be used in an evaluation of initial biases to the two halves of the alphabet. Load refers to the number of different categories per slide (4, 6, 8, 10, and 12), while Category Value refers to the two populations of categories which were designated as high and low value for each of the ratios.

Each S served for approximately 40 min. During the first 15 min., Ss received practice in seeing slides, in writing answers during the interslide interval, and in differentiating between the halves of the alphabet. All instructions for the experimental period were given immediately following practice. The E explained that points would be received for each letter reported correctly and that the number of points received for any particular letter was equal to the value assigned to the category population in which it was contained. For example, in an 8:1 ratio group, if the letters A–K were designated as the high-value categories and P–Z were the low-value categories, then an S would receive 8 points for each letter within A–K that was correctly reported and 1 point for each correct report of the letters P–Z. The only restriction placed on the report of letters was that points would be received only for each different category that was reported. This restriction was necessary since each particular letter on a slide was always repeated twice. The E further explained that there was no penalty or loss for incorrect or false reports. That is, S was instructed to report each detection, even if some uncertainty was present. Although Ss knew the values of the categories, they were given no knowledge of the load, of the frequency of occurrence of categories, or of results.

To eliminate possible biases in favor of the first or second half of the alphabet, the ratios of value were counterbalanced for the two halves. Thus four Ss within each group of eight had the first half of the alphabet as the higher value categories, while the reverse was true for the other four. All Ss were presented with all 50 slides in the same random order. A 10-sec. interslide interval was provided to permit a report of the letters which were detected and recalled. Following the experimental period, each S was asked to record the method employed to search the slide.

Subjects.—One hundred and twenty male volunteers, enrolled in the summer session at the University of Massachusetts, served as *S*s. Each of the eight *S*s within a group competed for a monetary award of $5.00 and two awards of $2.00.

Results

The data were analyzed in terms of three measures: (*a*) the percentage of high- and low-value categories which were correctly reported, (*b*) the percentage of times that the initial report was a high-value letter, and (*c*) the number of false reports of high- and low-value letters, i.e., the report of a letter when it was not displayed. The analyses of these measures were concerned with the effects of differential value and did not include the results for the 1:1 groups.

The results for the percentage correct for each condition were obtained from the combined number of correct reports over the 10 slides in each of the levels of category. An analysis of variance was performed on the arc sine of these data. The results of this analysis are summarized in Table 1 where it may be seen that the main effects of Exposure Time, Load, Category Value, and Ratio of Value were all significant. Of the first-order interactions, Load × Exposure Time, Category Value × Exposure Time, and Category Value × Ratio of Value were significant effects. The rest of the first-order interactions and all of the high-order interactions were not significant sources of variance. Summaries of the major significant effects are presented in Fig. 1, 2, and 3.

TABLE 1

Summary of Analysis of Variance for Accuracy of Report of High- and Low-Value Categories

Source	df	MS	F
Between *S*s	95		
Exposure time (E)	2	3624.25	318.62**
Ratio of value (R)	3	405.26	3.95*
E × R	6	209.85	2.05
*S*s/E × R	84	102.39	
Within *S*s	864		
Load (L)	4	26821.52	1431.09**
Category value (V)	1	8552.10	88.17**
L × V	4	53.90	1.91
L × E	8	1411.52	75.31**
L × R	12	8.00	.42
V × E	2	909.20	9.37**
V × R	3	318.83	3.28*
L × V × E	8	43.85	1.55
L × V × R	12	7.65	.27
L × E × R	24	11.50	.61
V × E × R	6	88.98	.91
L × V × E × R	24	21.31	.75
*S*s × L/E × R	336	18.74	
*S*s × V/E × R	84	96.99	
*S*s × L × V/E × R	336	28.15	

* $p < .05$.
** $p < .01$.

Figure 1 presents accuracy of performance as a function of load and exposure time. Inspection of this figure suggests that the amount of benefit derived from increasing exposure time was inversely related to the number of categories displayed, and that percentage correct varied inversely with the number of categories displayed (load) and directly with the length of exposure.

Figure 2 shows accuracy of performance as a function of ratio of value, with category value as a parameter. Although the 1:1 groups were not included in the analysis of variance, these data are included in the figure to illustrate the level of performance for equal-value categories. A comparison of the 1:1 condition with the rest of the groups shows the strong effect of the addition of differential category value. That is, the accuracy of report for the equal-value categories was always greater than for the low-value categories and less than for the high-value categories.

The significant main effect of Ratio of Value may also be inferred from Fig. 2. Although there was a slight increase in accuracy with the 2:1 ratio, performance decreased as the ratio was increased to 8:1. The increase in percentage correct at 16:1 did not bring performance back to the

level attained with the 1:1 and 2:1 ratios. Inspection of the Category Value × Ratio of Value interaction, omitting the 1:1 data, indicates that the decrement in overall performance was due to a reduction in accuracy for the lower value categories as the differential ratio of value increased. Thus, the figure suggests that the interaction may not represent an increase in accuracy of high-value letters, but a decrease in performance for the low-value letters. Evidence for this suggestion was obtained from the results of analyses of the simple main effects of ratio of value at each of the category value populations. For the high-value categories, increasing the ratio of value from 2:1 to 16:1 produced no significant difference in accuracy, while performance with the low-value categories was significantly affected, $F(3, 84) = 5.81$, $p < .01$. A Newman-Keuls test (Winer, 1962) of this significant effect revealed that mean performance for the low-value categories at 2:1 was superior to performance at 8:1 ($p < .01$) and 16:1 ($p < .05$) and also that a greater accuracy was attained with the 4:1 ratio than with the 8:1 ($p < .05$) ratio.

Figure 3 presents the interaction of Category Value with Exposure Time. It can be seen from this figure that the difference in percentage correct be-

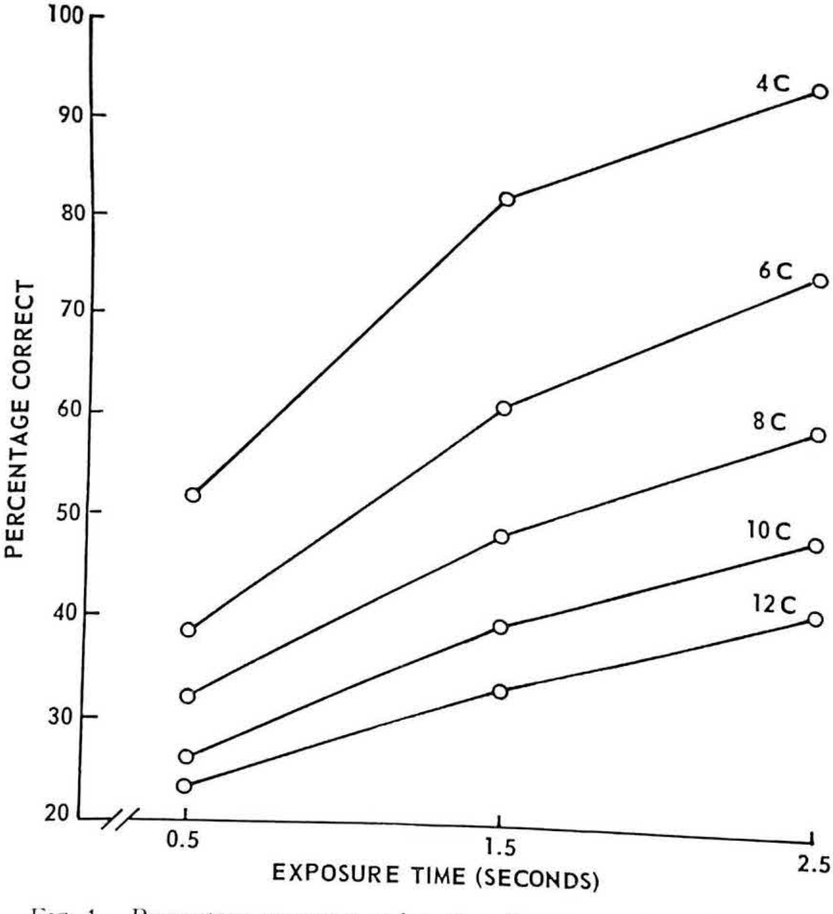


FIG. 1. Percentage correct as a function of exposure time with load (number of categories) as a parameter.

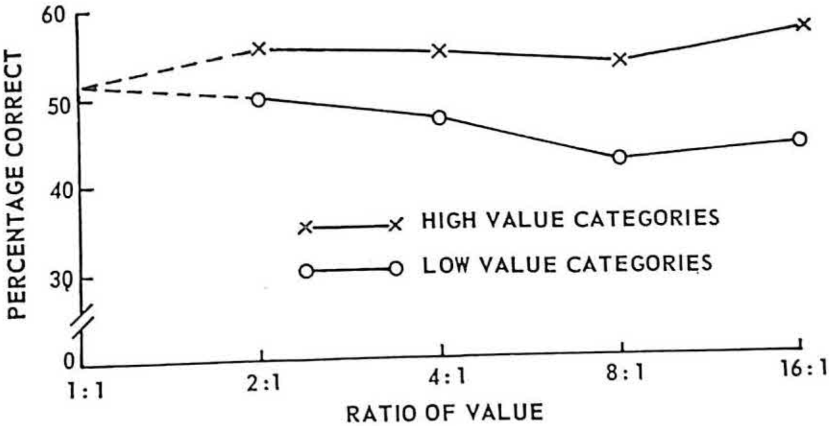


FIG. 2. Percentage correct of high- and low-value categories as a function of ratio of value.

tween high- and low-value categories increased directly with exposure time. The divergence of the lines suggests that an increase in exposure time gives greater benefit to performance with high-value categories than to performance with the low.

The initial response data were based on a total of 10 possible responses in each condition. Unlike total percentage correct, the data for the high- and low-value initial responses did not vary independently, i.e., the scores for the low-value responses were equal to the difference between 100% and the percentage of high-value responses made. An analysis of variance performed on the arc-sine transformation

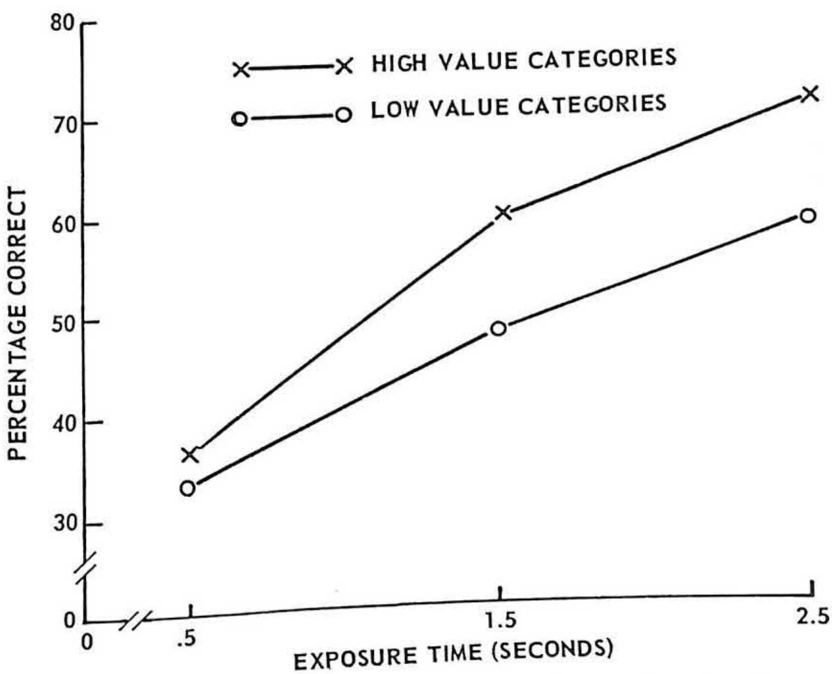


FIG. 3. Percentage correct of high- and low-value categories as a function of exposure time.

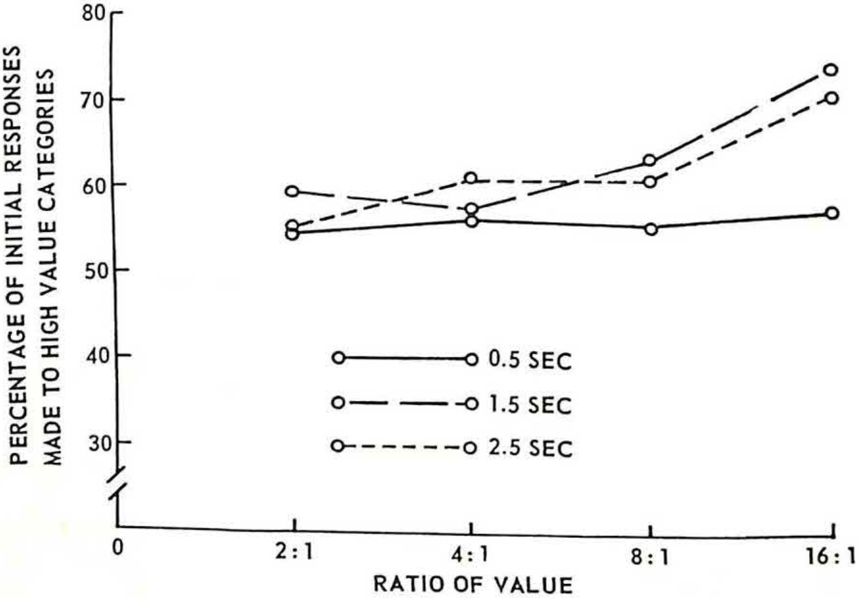


FIG. 4. Percentage of initial responses of high-value categories as a function of differential ratio of value with exposure time as a parameter.

of the initial responses to the high-value letters indicated significance for Exposure Time (E), $F\ (2, 84) = 3.31$, $p < .05$ and Ratio of Value (R), $F\ (3, 84) = 6.08$, $p < .01$. The main effect of Load and the interactions among E, R, and L were not significant sources of variance.

Figure 4 shows the percentage of initial responses of higher value letters as a function of ratio groups with exposure time as a parameter. Inspection of this figure indicates that all groups made initial responses of high-value categories more than 50%

TABLE 2

SUMMARY OF SEARCH METHODS EMPLOYED BY EACH *S* AT EACH EXPOSURE TIME

Exposure Time	Methods of Search		
	Restricted Area	Entire Screen	Scanning
0.5	26	6	8
1.5	8	3	29
2.5	3	0	37

of the time and that the percentage of these responses varied directly with the differential ratio of value. Figure 4 further suggests that the initial report of high-value letters was a maximum at 1.5 sec. exposure. A Newman-Keuls test showed that a significantly greater percentage of reports was made at 1.5 sec. than at 0.5 sec. ($p < .05$). The difference in performance between the means at 1.5 and 2.5 sec. and at 0.5 and 2.5 sec. was not significant. This suggests that 1.5 sec. may be either an optimum exposure time or an asymptote of higher value initial responses.

In general, the total number of false reports was relatively small in comparison to the number of correct reports or the number of categories omitted from reports. The mean number of false reports pooled over all 50 slides was 9.51 indicating that *S*s made approximately 1 false report for every 5 slides that they observed. An analysis of variance of the total number of high- and low-value false reports

summed over all 50 slides revealed that the only significant source of variance was Category Value, $F(1, 84) = 5.51$, $p < .05$. The mean number of false reports of high-value letters was 10.39, while the mean for the low-value categories was 8.64. The nonsignificant Category Value × Exposure Time and Category Value × Ratio of Value interactions suggest that the differential false report response to category value was not affected by increasing exposure time or ratio.

Table 2 presents a summary of the question concerning responses to the question concerning the method of search behavior. The "restricted area" responses refer to statements indicating an inability to search more than a small portion of the slide. "Entire screen" refers to those statements which indicated that *S* tried to fixate on the whole screen. Finally, "scanning" included those responses which indicated that *S* searched or attempted to search the whole screen in some systematic fashion. Inspection of Table 2 indicates that method of search was affected by the length of slide exposure. With an exposure time of 0.5 sec., most *S*s were able to search only a limited area of the slide, while at 1.5 and 2.5 sec., the majority of *S*s were able to systematically scan the slide. This suggests that the major effect of increasing exposure time was to allow *S*s to search a larger area and thus to detect more of the letters.

Although the ratio of values for the two halves of the alphabet were counterbalanced to eliminate the possible effects of preferences, two analyses were run to determine if any preferences did, in fact, exist. The dependent measures for these analyses were the percent correct of first and second half categories and the percentage of initial responses of categories within the first half of the alphabet. The 1:1 groups were included in these analyses since the objectives were to analyze the effects of category halves and not differential value. With percent correct as the dependent measure, the preferences for the two halves of the alphabet varied with both number of categories and exposure time, $F(8, 420) = 3.84$, $p < .01$. The increase in exposure time from 0.5 sec. to 2.5 sec. eliminated a first half preference with 4 and 6 categories and produced a second half preference for 8 and 10 categories. The slight preference for the second half of the alphabet with 12 categories did not change. In general, the results for the initial reports indicated first half preferences with 4, 6, and 8 categories, second half preferences with 10 and 12 categories, but revealed no consistent trends with exposure time.

Discussion

The results of those effects which did not include differential value are in complete agreement with previous studies (Teichner & Myers, 1961, 1962; Teichner et al., 1961; Teichner & Sadler, 1962). Further, the data suggest that differential value did have an effect upon *S*s' reports. Evidence for this effect comes from the fact that *S*s made more correct identifications, initial responses, and false reports to the higher value categories than to those of lesser value.

Although the results supported the hypothesis of differences in accuracy due to differential value, the ratio of value effect and the exact shape of the category value by ratio of value interaction were not expected. The findings of previous studies indicate that the number of correctly reported stimuli remains invariant over a wide range of conditions (Sperling, 1960) and that percentage correct remains fairly constant for any particular combination of exposure, load, and density (Teichner & Myers, 1961, 1962).

Using these studies as the basis for prediction, it was assumed that overall accuracy would not vary as ratio of value was increased, and that every increase in accuracy for the higher value categories would be accompanied by a corresponding decrease in performance for the lower value categories. The results indicate that although there was a decrement in performance for the lower value letters, there was no corresponding increment for the higher value categories. This loss in accuracy without any compensating increment resulted in a decrement in overall performance as the ratio of value was increased. Thus, the data suggest that an increase in differential ratio does not seem to facilitate reports of the more important targets, but rather inhibits the reports of ones of lower value.

The only dependent measure which was not affected by increasing the differential ratio of value and exposure time was the number of false reports. Since the present experiment provided an incentive for the higher value categories and no penalties for false reports, the results were contrary to predictions based on the theory of visual signal detection (Tanner & Swets, 1954). On the other hand, the data agree with a study that found that a set to guess did not affect the number of letters reported in the visual search task (Teichner & Myers, 1962). This discrepancy in findings suggests either that signal detection theory does not apply to the present search task in which each letter was clearly discriminable, or that the differences were caused by variations in response criteria. With respect to the response criteria, the present study provided no information concerning the number of categories to be presented and placed no restrictions on the number of responses, while the study of signal detection used forced-choice responses. It is possible that if a specified number of reports was required in the search task, then the number of false reports would vary with differential ratio of value.

Studies by Teichner, Reilly, and Sadler (1961) and Sperling (1960) have demonstrated that there is a difference between the amount of information received and stored, and the number of letters which can be recalled. Teichner and Myers (1961, 1962) assumed that every letter that is perceived enters into short-term memory storage and that recall represents the difference between storage and memory loss due to some process going on in the storage. This assumption implies that differential value may have affected either perception or recall. Since the present experiment used recall as the measure of performance, the effects of selective perception of the categories and of selective recall from memory storage were not separable and the relative importance of these factors could not be evaluated.

On the other hand, if a major selective process in perception was involved, then the size of the difference between high- and low-value letters should vary inversely with length of exposure. That is, a perceptual set, which would make the higher value categories easier and quicker to identify (Bruner, 1957), would occur at those conditions where there is little time to detect and discriminate between the categories, but have little effect at longer exposures where all letters would be equally discriminable. Predictions from a selective recall point of view would be the opposite. That is, as exposure time increases and the rate of memory loading decreases, *S*s would have more time to selectively code the categories which are detected. Thus, a selective memory process explanation would predict increasing differences in response to value as exposure time increases. Since the results support this latter hypothesis, the evidence is in favor of a selective process between memory storage and recall. The explanation in terms of a perceptual bias would also predict that frequency of false reports would vary directly with differential ratio of value and inversely with exposure time. Although significantly more false reports were made to the higher value categories, the small number of reports and the lack of significant changes with exposure time and ratio of value suggest that selective perception

was not a major factor in the present study. Further, the verbal reports of *S*s suggest that lengthening slide exposure did not seem to aid perception, but only allowed more letters to be detected through increased scanning.

Finally, the choice and arrangement of the stimulus materials were designed to minimize the effects of a selective perception. All letters were of equal size and were clearly visible and discriminable; and each slide contained an equal number of randomly selected and randomly positioned letters from the two halves of the alphabet. Thus, it is assumed that an equal number of high- and low-value categories was detected and that the differential effects were due to a selective decision which occurred after the development of memory storage. This selective process is assumed to result in a coding of inputs and subsequent retrieval in favor of the more important categories.

Although the results support the assumption of a selective recall from memory storage, other findings in favor of selective encoding in perception (Harris & Haber, 1963) indicate a need for research to evaluate the relative importance of each of the hypothesized selective processes. In order to adequately assess the selective aspects of perception and memory, provision must be made, within a single task, to separate, control, and measure the amount and complexity of information which is perceived and stored and the amount of information which is retrieved and recalled. This could be accomplished by adapting the method utilized by Teichner, Reilly, and Sadler (1961) for the search task, or by presenting the stimuli sequentially. Further, research is necessary to determine the effects upon performance with reward for correct responses and penalties for errors of omission and commission. Finally, research on response criteria should be performed to relate the present findings to signal detection theory.

REFERENCES

Anderson, N. S., & Fitts, P. M. Amount of information gained during brief exposures of numerals and colors. *J. exp. Psychol.*, 1958, **56**, 363–369.

Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–150.

Harris, C. S., & Haber, R. N. Selective attention and coding in visual perception. *J. exp. Psychol.*, 1963, **65**, 328–333.

Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

Tanner, W. P., & Swets, J. A. A decision making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

Taub, H. A., & Myers, J. L. Differential monetary gains in a two-choice situation. *J. exp. Psychol.*, 1961, **61**, 157–162.

Taub, H. A., & Teichner, W. H. Effects of differential value and exposure time upon the detection and memory of symbols in a visual search task. *USAF ESD tech. docum. Rep.*, 1963, No. 63-343.

Teichner, W. H., & Myers, J. L. Training aspects of decision making. *USN Train. Devices Cent. tech. Rep.*, 1961, No. 588-1.

Teichner, W. H., & Myers, J. L. Training aspects of decision making, phase II. *USN Train. Devices Cent. tech. Rep.*, 1962, No. 588-2.

Teichner, W. H., Reilly, R., & Sadler, E. Effects of density on identification and discrimination in visual symbol perception. *J. exp. Psychol.*, 1961, **61**, 494–500.

Teichner, W. H., & Sadler, E. Effects of exposure time and density on visual symbol identification. *J. exp. Psychol.*, 1962, **63**, 376–380.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received December 26, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 144–151

CONCEPT LEARNING AND VERBAL CONTROL UNDER PARTIAL REINFORCEMENT AND SUBSEQUENT REVERSAL OR NONREVERSAL SHIFTS [1]

DANIEL C. O'CONNELL [2]

University of Illinois

Concept learning was studied with partial or continuous reinforcement, reversal or nonreversal shift, and reinforcement for placement or reason. Ss placed cards right or left and gave trial-by-trial reasons; half were reinforced for placements, half for reasons. In acquisition, half received 100%, half 60% reinforcement, until each reached 50 reinforcements. Shift period was 80 trials; half the Ss were reinforced for reversal, half for nonreversal shift. In acquisition and shift period, PRE was found. Reinforcement for placement or reason differed significantly. Reversal or nonreversal shift gave inconclusive evidence concerning mediation theory. A theory of verbal control accurately predicted correct placements from verbalized reasons. Number of nonreinforced acquisition trials correlated negligibly with shift-period response measures.

Investigations concerned with partial reinforcement of the stimulus dimension relevant during acquisition of a concept by human *S*s have yielded rather complex results. Evidence regarding the existence of a partial-reinforcement effect (PRE) remains in several respects negative or inconclusive.

Buss (1952) found *less* resistance to extinction after partial reinforcement (50%) than after continuous. He noted that the experiment differed from a standard extinction experiment in that the response was attached to a new stimulus while being extinguished on the first. Still he could find no obvious explanation for the prolonged extinction after continuous reinforcement.

Sax (1960) utilized the same percentages of reinforcement as Buss and four delays of reinforcement. His measure of habit strength at the end of acquisition was the number of trials needed to reach a criterion of perfect performance. Neither schedule of reinforcement nor its interaction with delay of reinforcement proved to be significantly related to this measure. Retention after 2 wk. rather than a standard extinction procedure was the second measure of habit strength. Schedule of reinforcement was found to be unrelated to this measure also.

Erlebacher and Archer (1961) found that, as percentage of reinforcement increased (25%, 50%, and 100%) during acquisition, number of correct responses, number of errors, and total number of responses decreased. Their *S*s were required to reach a degree of learning defined by a criterion of 10, 20, or 40 consecutive correct responses. They considered a nonreversal shift analogous to an extinction condition

[1] This article is based on a dissertation submitted to the Department of Psychology of the Graduate College of the University of Illinois in partial fulfillment of the requirements for the PhD degree. The author is especially indebted to Jack A. Adams for his guidance and assistance through the course of this investigation. Supplementary experimentation reported in the article has been supported by a National Science Foundation Postdoctoral Fellowship.

[2] Now at the Center for Cognitive Studies, Harvard University.

and hypothesized accordingly that number of perseverative errors would be inversely related to percentage of reinforcement. This second aspect of PRE was not confirmed, although there did appear a significant interaction between percentage of reinforcement and degree of learning, with perseverative errors as the response measure. During acquisition, however, increase in percentage of reinforcement was confounded with a decrease in *number* of reinforcements—a problem in partial-reinforcement studies to which Kanfer (1954) has called attention. In the present instance, it renders the negative results and the already complex interaction difficult to interpret.

The analogy of reversal and nonreversal shifts to an extinction condition makes the relative difficulty of the two types of shift relevant. Kendler (1960) has reviewed the evidence that the majority of adult *S*s learn a reversal shift rather than a nonreversal shift. The former requires a shift only in the overt response and not in the postulated mediating response, and this accounts for its relative facility. Mediation theory also predicts positive transfer in a reversal shift as compared with a control, but Isaacs and Duncan (1962) have obtained negative results.

It is also of interest to inquire whether PRE can be found when reinforcement is applied to verbal rather than motor responses. A design which can be adapted to test such a hypothesis has been presented by Verplanck (1962). He found evidence of the dissociation of motor (card placements) and verbal (trial-by-trial hypotheses) responses under partial reinforcement (60%) after acquisition under continuous reinforcement. Dulany (1962), on the other hand, has presented a theory of verbal control which predicts no dissociation under such conditions, and Dulany and O'Connell (1963) have presented evidence to the same effect from a replication of Verplanck's research.

In both the Sax (1960) and Erlebacher and Archer (1961) studies, the failure to control for the number of reinforcements made the meaning of the negative results regarding PRE in the second phase of the experiments unclear. Moreover, Sax used a retention measure, and Erlebacher and Archer used only a nonreversal shift. In this study, it is expected that PRE will be verified in both acquisition and shift period, with reinforcement contingent on motor or verbal responses. The negative findings of Isaacs and Duncan (1962) suggest the further hypothesis that reversal and nonreversal shifts will not prove significantly different. Finally, dissociation of placement and reason is expected in none of the conditions in consequence of the various reinforcement contingencies.

Method

Stimuli.—A set of 320 3 × 5 in. cards was used. The figures thereon differed along four dimensions: shape (20 different kinds), position (top or bottom of the card), color (red, black, blue, or green), and number (one or two figures per card). Order of presentation was the same for all *S*s; each block of 20 cards had one card with each of the 20 figures, 5 of each color, 10 with figures in the top and bottom positions, respectively, and 10 with one and two figures, respectively. Within each block, order of presentation was randomized.

One block of 20 cards served as a sample for *S*s during instructions, and an extra card covered the deck. The other 300 cards were piled face up in order before *S*. To the left and right were receptacles for the cards.

Subjects.—The sample consisted of students in an introductory psychology course. The 20 *S*s in each of the eight experimental conditions were randomly assigned.

Procedure.—Each *S* was scheduled individually for an hour. He was seated at a table with *E* at his side and somewhat behind him. A screen hid *E*'s data sheets. Instructions were as follows:

> The experiment you're helping me with is essentially a card sorting task; the cards can be sorted systematically into two stacks in a number of ways. You are simply to take each card in turn, decide whether to place it on the right or left stack, and give *the reason* why the card should be so placed. Tell me your reason in the following way: "— figures go to the right (left)." Meanwhile place the card accordingly and take the next card. The blank space is to be filled in with *a single word* descriptive of the card. Your reason might be, for example, "*Green* figures go to the left," or "*Triangular* figures go to the right." What the real reasons are you must find out for yourself in the course of the experiment. I will reply with either "yes" or silence after each trial. *You are to try to get as many "yeses" as possible.* Here are some sample cards from the deck. Look through them to familiarize yourself with the type of card you will be dealing with during the experiment. Now to make sure you understand how the cards can be sorted into two piles, notice that the samples include *green* figures, *red* figures, *black* figures, and *blue* figures; *single* figures and *double* figures; *top* figures and *bottom* figures; and a number of *types* of figures, or *shapes*. These are the dimensions you are to use in filling in the blank space in the formula when you give your reasons.

The *S*'s questions were then answered, examples of all possible logical reasons were given with the extra card, and *S* was reminded of the nature of his task. He then proceeded through the cards, pacing himself and without interruption, and filled out a brief post-experimental questionnaire.

Conditions.—Throughout the experiment, the P group (80 *S*s) were reinforced ("yes" from *E*) for correct placements irrespective of their reasons; the H group (80 *S*s) were reinforced only if the reason for their placement was exactly correct. During acquisition, each group was subdivided into groups with 100% and 60% reinforcement, respectively. In the latter group, *E*'s "yeses" were given in a randomized three out of each series of five correct responses (placements or reasons). Acquisition lasted until *S* had been reinforced 50 times. Then all groups were shifted without notice to 100% reinforcement and subdivided into reversal- and nonreversal-shift groups for 80 more trials.

During acquisition, correct placements were: cards with one figure to the right; cards with two figures to the left. During the shift period, correct placements for the reversal-shift group were: cards with one figure to the left; cards with two figures to the right. For the nonreversal-shift group, they were: cards with figures at the top to the right; cards with figures at the bottom to the left.

Supplementary conditions.—An additional 30 *S*s, volunteers from Harvard and Radcliffe Colleges, were assigned to several of the experimental conditions in order to obtain further evidence regarding verbal control. From the instructions above, the sentence requiring that *a single word* be used, the sentence beginning "What the real reasons are," and the last two sentences were deleted. Twenty *S*s were assigned to the P group and evenly subdivided into groups with 100% and 60% reinforcement, respectively. The group with 100% reinforcement was given acquisition and nonreversal-shift trials; the group with 60% reinforcement was given acquisition trials only. The remaining 10 *S*s were assigned to the H group and given acquisition trials with 60% reinforcement.

RESULTS

Learning during acquisition.—A number of response measures indicated terminal level of learning in each of the four acquisition groups: number of *S*s whose last 20 placements were all correct; number of *S*s whose last 20 reasons were all correct; mean number of correct placements in the last 20 trials for all *S*s; and mean number of correct reasons in the last 20 trials for all *S*s. For all these response measures, the differences between the 100% and 60% reinforcement conditions were significant in both P and H groups.

The binomial test was applied to the differences in number of *S*s whose last 20 placements were all correct: for the H groups (39 and 14), $p < .001$; for the P groups (26 and 2), $p < .001$. The binomial test was similarly applied to the differences in number of *S*s whose last 20 reasons were all

TABLE 1

Mean Response Measures during Shift Period

Group	N	Placements Correct	Reasons Correct	Perseverative Reason Errors	Perseverative Placement Errors
H-100-R	20	73.00	69.20	3.15	7.00
H-100-N	20	74.70	71.25	3.10	3.05
H-60-R	20	55.05	51.80	20.55	24.95
H-60-N	20	55.35	35.20	26.40	18.25
P-100-R	20	67.20	54.25	3.75	12.80
P-100-N	20	63.90	50.35	7.70	8.30
P-60-R	20	38.40	9.55	10.10	41.60
P-60-N	20	49.60	19.75	18.40	22.60

correct: for the H groups (35 and 9), $p < .001$; for the P groups (23 and 1), $p < .001$.

The t test was applied to the mean number of correct placements in the last 20 trials for all *S*s: for the H groups (19.95 and 17.95), t (38) = 4.47, $p < .01$; for the P groups (18.02 and 14.52), t (38) = 4.32, $p < .01$. The t test was similarly applied to the mean number of correct reasons in the last 20 trials for all *S*s: for the H groups (19.72 and 15.82), t (38) = 5.93, $p < .01$; for the P groups (14.95 and 5.95), t (38) = 6.55, $p < .01$.

Learning during shift period.—The main response measures for the shift period were: correct placements, correct reasons, perseverative reason errors, and perseverative placement errors. Means are presented summarily in Table 1.

A $2 \times 2 \times 2$ factorial analysis of variance for correct placements indicated two significant main treatment effects: for the PRE, F (1, 152) = 93.09, $p < .01$; for the H-P effect, F (1, 152) = 16.14, $p < .01$. Parallel results were found in the analysis for correct reasons: for the PRE, F (1, 152) = 77.03, $p < .01$; for the H-P effect, F (1, 152) = 40.67, $p < .01$. In both these analyses, Bartlett's test indicated heterogeneity of variance: χ^2 (7) = 26.18, $p < .01$; and χ^2 (7) = 27.82, $p < .01$. Hence the level accepted as significant was .01 in both instances (Lindquist, 1953).

In the analysis for perseverative reason errors, however, a transformation (log of score plus one) was effective in eliminating heterogeneity. In this analysis, the significant results were: the PRE, F (1, 152) = 69.97, $p < .01$; the H-P effect, F (1, 152) = 4.29, $p < .05$; the reversal-nonreversal effect, F (1, 152) = 5.83, $p < .05$; and the triple interaction, F (1, 152) = 18.10, $p < .01$.

The experimental design allowed only half as many perseverative placement errors in nonreversal as in reversal conditions. Hence a direct comparison of these conditions was not feasible, and two separate 2×2 factorial analyses were required. In the analysis for reversal conditions, the two main treatment effects were significant: for the PRE, F (1, 76) = 52.26, $p < .01$; and for the H-P effect, F (1, 76) = 12.05, $p < .01$. In the analysis for nonreversal conditions, only one main treatment effect was significant: for the PRE, F (1, 76) = 63.27, $p < .01$. In the analyses for both reversal and nonreversal conditions, Bartlett's test indicated heterogeneity of variance: χ^2 (3) = 13.85, $p < .01$; and χ^2 (3) = 20.61, $p < .01$. Again, the level accepted as significant was .01 because of the heterogeneity.

Similar comparisons were made by applying the binomial test to the number of *S*s in the various conditions to reach 10 or more successive correct placements or correct reasons. In terms of both criteria, the PRE was significant (73 and 40; 69 and 34), $p < .01$, as was also the H-P effect (70 and 43; 69 and 34), $p < .01$. The frequencies for the reversal-nonreversal effect (56 and 57; 55 and 48) were not significant.

The number of *S*s whose first shift in reason during the shift period was a reversal was significantly different in the H and P conditions (38 and 11) by the binomial test, $p < .001$. This result is hardly surprising, since learning of reasons during acquisition reached a higher level in the H than in the P conditions. The number of *S*s in the H condition whose first shift in reason during the shift period was a reversal and nonreversal, respectively (38 and 41) was then compared by means of a 2×2 contingency table with the frequencies for these same reasons in the initial responses of the experiment (41 and 86). The shift-period frequencies were found to be significantly different from the expected frequencies: $\chi^2\,(1) = 16.86$, $p < .001$.

Verbal control.—Miscategorizations (*S* named a figure erroneously, e.g., called a green figure red) occurred on only 6, or .02% of the 29,874 trials of the experiment. Misplacements (*S* placed card on one side while saying he was placing it on the other side) occurred on 58, or .19% of the trials. Misplacements for the 100% and 60% reinforcement conditions during acquisition were 25 and 10, respectively. However, the larger figure was contributed to by multiple misplacements of several *S*s. The cumulative distributions of *S*s making various numbers of misplacements in these conditions were compared by means of the Kolmogorov-Smirnov two-sample test, and the differences proved nonsignificant: $D = .0875$, $p > .10$.

A formula for expected correct placements (P_E) was adapted from Dulany and O'Connell (1963). It relies on the following assumptions: Correct reasons control correct placements; Incorrect reasons control incorrect placements; Irrelevant reasons control correct and incorrect placements with equal frequency. Correlated reasons (e.g., reasons correct but for the addition of superfluous elements) had proved an important component of the original formula, but were excluded from the adapted formula because *S*s were free to use only a single word descriptive of the card in their reasons. This constraint was introduced in order to limit the experimental time for each *S* to 1 hr. and avoid nonlearners. Without such a constraint, Verplanck (1962) and Dulany and O'Connell (1963) had found it necessary to discard 17 and 25 *S*s, respectively.

Correct and incorrect numerical reasons were coded as H_N and H_n, respectively; correct and incorrect position reasons as H_P and H_p, respectively. Irrelevant color and shape reasons were coded as H_C and H_S, respectively. For conditions in which the correct reason was numerical, the adapted formula was:

$$P_E = H_N + .5(H_C + H_S + H_P + H_p)$$

For conditions in which the correct reason involved position, the adapted formula was:

$$P_E = H_P + .5(H_N + H_n + H_C + H_S)$$

The *t* test for related measures was applied to the differences between expected and observed placements for each of the four acquisition and eight

shift-period conditions. In three acquisition conditions, the differences proved significant: for the 60% H condition, $t\ (39) = 3.62$, $p < .001$; for the 100% P condition, $t\ (39) = 3.68$, $p < .001$; and for the 60% P condition, $t\ (39) = 4.70$, $p < .001$. One of the shift-period conditions proved significant: for the nonreversal 100% P condition, $t\ (19) = 2.11$, $p < .05$.

Nonreinforced trials during acquisition.—For each condition, correlations were computed between number of nonreinforced trials during acquisition and each response measure of the shift period: correct placements, correct reasons, perseverative reason errors, and perseverative placement errors. Of the 32 correlations, 7 proved significant, and of these, 2 were positive, 5 negative. They were distributed over conditions with no discernible regularity, even in the case of perseverative errors, for which the analogy to a standard extinction condition is most relevant.

Supplementary conditions.—For these data, the *original* formula for expected correct placements (Dulany & O'Connell, 1963) was applicable. Lifting the restriction to a single word in filling in the blank space made it necessary to discard 10 nonlearners; but it also made it possible for *S* to verbalize reasons which, although too specific to be scored as logically correct, could be expected to have a perfect positive correlation with correct placements. The *t* test for related measures was applied to the differences between expected and observed placements for the three supplementary acquisition conditions and the one supplementary nonreversal-shift condition. These were the very conditions for which the *t* tests proved significant in the results of the main experiment, and yet none of the *t* tests in the present instance even approached significance.

Discussion

Partial-reinforcement effect. — The present results clearly indicate PRE in both acquisition and shift period. Every one of the placement and reason response measures indicates the effect at the .01 or .001 level of significance. During acquisition, terminal level of training is consistently higher under continuous than under partial reinforcement. During the shift period, resistance to extinction is consistently higher in the group which had partial reinforcement during acquisition than in the group which had continuous reinforcement during acquisition.

Lawrence and Festinger (1962) have recently rejected the explanation of PRE in terms of contrast between acquisition and extinction trials and explained the phenomenon as a function of the total number of unrewarded trials. In the results of the present experiment, however, there are no significnat correlations between number of nonreinforced trials during acquisition and any response measure for the shift period which disclose such a systematic or consistent relationship. The few correlations which did prove significant reflect the trivial fact that individual *S*s are consistently good or poor in their performance during acquisition and the shift period. It should be noted, however, that the similarity of shift periods to an extinction condition is analogous; the evidence found in this experiment cannot be extended to a standard extinction condition in every respect without further evidence.

Reversal and nonreversal shifts.—The evidence of the present experiment does not lend support to the hypothesis that a reversal shift will yield faster or better learning than a nonreversal shift. Only the analysis for perseverative reason errors yields a significant reversal-nonreversal effect, but a study of the means in Table 1 shows that the largest reversal-nonreversal differences appear in the partial-reinforcement groups. Duncan's

multiple-range test shows that these are also the differences which account for the significant effect. In terms of the mediation hypothesis, this would mean that, where the mediator was learned poorly during acquisition, reversal and nonreversal shifts differentiate subsequent levels of learning the better; where the mediator was learned well, there appear no significant differences between the reversal and nonreversal groups. Even this isolated significant result is, therefore, exactly the opposite of what the mediation theory would hypothesize.

It is plausible to suggest that the failure to find significant differences between reversal and nonreversal shifts might be a consequence of the different salience of the dimensions relevant in the two stages of the experiment. Since these dimensions were not varied, they were indeed confounded with the shift variable; the relevant dimensions for reversal and nonreversal shifts were always numerical and positional, respectively. If position were a more salient dimension than number, the nonsignificant result could be expected. But the initial responses of the experiment indicate no such salience. On the contrary, 74 *S*s gave numerical reasons as their first response, and only 19 gave positional reasons. Hence, the salience of the relevant dimensions would have inflated any difference in favor of reversal shift over nonreversal shift.

The evidence does support the mediation hypothesis, however, insofar as the number of *S*s in the H conditio i whose first shift in reason was a reversal was significantly greater than expected in accord with the salience of these same reasons in the initial responses of the experiment. Apart from the salience of dimensions, we should expect only 11 reversal-shift responses rather than the 38 observed in the 79 first shifts in reason, if a reversal shift occurred only as frequently as each of the other six available (nonreversal) shifts. The preponderance of reversals among these first shifts in reason is, therefore, all the more impressive.

Verbal control.—Dulany and O'Connell (1963) found that miscategorizations accounted for what appeared to be dissociation in Verplanck's research. And the miscategorizations were due in turn to ambiguous stimuli. No miscategorizations occurred in their control experiment with unambiguous stimuli. In the present experiment, miscategorizations and misplacements do occur, but so minimally as to have little effect on the accuracy of the formula for expected placements.

But the *adapted* formula does not accurately predict in all cases. The failures in prediction can well be ascribed to the inadequacy of the formula and the corresponding constraints placed upon the experimental design, rather than to any weakness inherent in a theory of verbal control as such. This suggestion is convincingly borne out by the results of the supplementary conditions. When *S*s were allowed to verbalize correlated reasons, they did so on 27% of the trials. And when these correlated hypotheses were in turn taken into account in the *original* formula for expected correct placements, all the significant differences between expected and obtained correct placements disappeared. It is clear that the adapted formula did not provide adequate evidence of verbal control precisely because *S*s could entertain hypotheses which they were not allowed to verbalize. The constraints placed on verbalization served only to preclude the possibility of gathering adequate evidence regarding verbal control.

In both acquisition and shift period, therefore, with reinforcement contingent on either verbal or motor responses, PRE can be found in concept learning. No dissociation due to reinforcement contingencies was found, and evidence regarding the relative facility of reversal and nonreversal shifts was mixed.

REFERENCES

Buss, A. H. Some determinants of rigidity in discrimination-reversal learning. *J. exp. Psychol.*, 1952, **44**, 222–227.

Dulany, D. E., Jr. The place of hypotheses and intentions: An analysis of verbal con-

trol in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness: A symposium of research and interpretation.* Durham, N. C.: Duke Univer. Press, 1962. Pp. 102–129.

DULANY, D. E., JR., & O'CONNELL, D. C. Does partial reinforcement dissociate verbal rules and the behavior they might be presumed to control? *J. verbal Learn. verbal Behav.*, 1963, **2**, 361–372.

ERLEBACHER, A., & ARCHER, E. J. Perseveration as a function of degree of learning and percentage of reinforcement in card sorting. *J. exp. Psychol.*, 1961, **62**, 510–517.

ISAACS, I. E., & DUNCAN, C. P. Reversal and nonreversal shifts within and between dimensions in concept formation. *J. exp. Psychol.*, 1962, **64**, 580–585.

KANFER, F. H. The effect of partial reinforcement on acquisition and extinction of a class of verbal responses. *J. exp. Psychol.*, 1954, **48**, 424–432

KENDLER, T. S. Learning, development, and thinking. *Ann. N. Y. Acad. Sci.*, 1960, **91**, 52–63.

LAWRENCE, D. H., & FESTINGER, L. *Deterrents and reinforcement: The psychology of insufficient reward.* Stanford, Calif.: Stanford Univer. Press, 1962.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

SAX, G. Concept acquisition as a function of differing schedules and delays of reinforcement. *J. educ. Psychol.*, 1960, **51**, 32–35.

VERPLANCK, W. S. Unaware of where's awareness: Some verbal operants—notates, monents, and notants. In C. W. Eriksen (Ed.), *Behavior and awareness: A symposium of research and interpretation.* Durham, N. C.: Duke Univer. Press, 1962. Pp. 130–158.

(Received December 27, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 152–157

EFFECT OF INTERTRIAL ACTIVITY ON THE RELATIONSHIP BETWEEN AWARENESS AND VERBAL OPERANT CONDITIONING [1]

PAUL W. DIXON AND WILLIAM F. OAKES
University of Hawaii

4 groups of 25 Ss constructed sentences using a verb and 1 of 6 pronouns presented on an index card in a series of 6 blocks of 20 trials. In a 2 × 2 design, 2 of the groups, RN and RC, were reinforced for sentences begun with "I" or "we," while the other 2 groups, UN and UC were not reinforced. Groups RN and UN named colors between trials, while RC and UC had no such intertrial activity. After the conditioning trials an "awareness" interview was conducted. The reinforced groups increased significantly in number of "I-we" sentences constructed and did not differ in the amount of conditioning. The correlation usually found between "awareness" and conditioning scores was found in Group RC, but a nonsignificant correlation in the opposite direction was found for Group RN. (The correlations for RN and RC were significantly different.) It was concluded that the intertrial activity (color naming) interfered with the "awareness"-conditioning relationship, but not with the conditioning, per se, thus suggesting an automatic strengthening effect of the reinforcement, not mediated by cognitive processes, i.e., "awareness." The results were interpreted as not supporting the Dulany-Spielberger position on the relationship between "awareness" and conditioning, but rather as supporting the Verplanck-Krasner position.

A question which has arisen from the research on verbal operant conditioning (VOC) has to do with the role of "awareness," usually indicated by retrospective verbal report, in such learning. The early investigators in this area (e.g., Greenspoon, 1955; Taffel, 1955) eliminated *S*s from the data analysis whose interview results indicated "awareness" (of a response-reinforcer contingency), and thus implied that this learning had occurred without awareness. Verplanck (1962) and Krasner (1962) have held that there is no *necessary* relationship between awareness and VOC, i.e., that the effect of the reinforcer is direct, and not necessarily mediated by cognitive processes.

[1] Although a different analysis was made, the data reported herein were obtained by the first author for his MA thesis under the supervision of the second author. This report was prepared by the second author.

Other investigators (e.g., Dulany, 1961, 1962; Spielberger, 1962), primarily on the basis of repeated findings of a positive relationship between degree of "awareness" and amount of VOC for individual *S*s, have questioned the conclusion of the early VOC investigators, and, in contrast to the Verplanck-Krasner position, suggested that VOC *is* necessarily mediated by cognitive processes, i.e., "awareness" of a response-reinforcement contingency (e.g., Spielberger, 1962, p. 95).

Dulany (1962) has presented a theoretical network which breaks down what others call "awareness" (which he conceives of as a self-instructional set mediating the response in VOC) into three concepts: Reinforcement Hypothesis (RH), Behavioral Hypothesis (BH), and Behavior Intention (BI). RH, BH, and BI are all

assessed by means of retrospective verbal report.

Dulany (1962) presents results indicating that the higher the level of RH, BH, and BI reported by *S*, the greater the conditioning of "I-we" sentences, i.e., in the Taffel (1955) VOC procedure. In that same paper Dulany (1962) reported supporting evidence for his theoretical assumption that RH and BH should affect performance only through BI, viz., that the correlation between performance and BI, BH, and RH fell off in the expected direction—.71, .57, .36, respectively, each significantly less than the preceding.

The present experiment is designed to test certain assumptions in and predictions from the Dulany (1962) theoretical system. It is the feeling of the writers that the question of whether *S*'s increase in rate of emission of the reinforced class in the VOC situation is mediated by cognitive processes ("awareness") or not is still not settled, in spite of the Dulany and O'Connell tour de force ("Does Partial Reinforcement Dissociate Verbal Rules and the Behavior They Might Be Presumed To Control"),[2] in which they showed that the Verplanck (1962) results which cannot be accounted for within the Dulany (1962) theoretical network are probably artifactual. It seems likely to the writers that the typical finding of a positive relationship between "awareness" and VOC may be primarily due to the fact that such a simple and obvious contingency has been involved. The Taffel (1955) procedure, which Dulany (1962) used, requires that *S* construct sentences that begin with one of several pronouns presented on a card, and *E* reinforces whenever *S* begins the sentence with "I" or "we." The *S* need not be too perceptive to detect this contingency between *E*'s reinforcer and *S*'s response in that situation, provided he is given the opportunity to formulate hypotheses regarding the contingency during the experiment.

The rationale for this experiment involves using that same Taffel (1955) procedure, but interfering with *S*'s opportunity to formulate such hypotheses between trials during the experimental session. Between trials some of the *S*s are required to perform a color-naming task, which would presumably interfere with such hypothesis-formulating behavior. If the Spielberger-Dulany position is correct, it would be expected that such interference between trials would reduce the possibility of conditioning in this situation, whereas if the Verplanck-Krasner direct-strengthening interpretation is correct, there should be no effect on the results with respect to conditioning, but for *S*s performing the color-naming task between trials the positive relationship usually found between "awareness" and conditioning scores should not be found. Further, RH, BH, and BI should not be related to conditioning scores in the way that Dulany found them to be.

Method

Subjects.—The *S*s were 100 students from introductory psychology classes at the University of Hawaii. They were randomly assigned to four groups of 25 each: two experimental groups, which both received reinforcement, one group with the color-naming task between trials, the other without; and two control groups, neither of which was reinforced, one with the color-naming task between trials, the other without.

Apparatus.—The stimulus materials for the conditioning task consisted of 3 × 5 in. white index cards on each of which a different neutral past tense verb was typed. These verbs were taken from the list used by Binder and Salop (1961). The pronouns I, WE, YOU, HE, SHE, and THEY were typed above the verb

[2] D. E. Dulany and D. C. O'Connell, personal communication, May 1963.

with the order of appearance of the pronouns randomized over all cards. The *S* was presented with easily recognizable colors shown in the aperture of a memory drum. The colors were arranged in random order by selection from a large assortment of colors.

A screen with its leaves folded in at approximately 120° toward *S* was used to separate *S* from *E*. The central portion of the screen was 26 × 30 in. and the leaves were 12 × 30 in. A place for the presentation of the card was made 12 in. above the base of the screen and in the center of the central panel. Beneath the hole through which the card was presented, a slot was cut in the screen to fit the memory drum so that the memory drum would face *S*. The colors on the memory drum thus appeared directly below the card holder.

The *E* used a scoring sheet on which the numbers 1–120 were followed by the pronouns used in the experiment, and he encircled the single pronoun used by *S* on each trial on that data sheet.

Procedure.—Although the results of the testing were not used in the present study, all *S*s were first administered the MMPI in one room before being brought to the experimental room. In the experimental room a short informal preconditioning interview was held with *E* in order to establish a friendly relationship between *E* and *S*. After this informal interview, the usual instructions for the Taffel (1955) procedure were read to *S*. The *S*s in the color-naming group were instructed to name the colors appearing on the memory drum between trials.

The *S*s then proceeded to construct sentences from the 120 cards on which the pronouns and the verb had been typed. For the reinforced groups, RC (reinforced control) and RN (reinforced naming), during the first 20 trials every third sentence constructed by *S* was followed by *E* murmuring "mm-hmm" regardless of which pronoun was used. This was done so that the first reinforcement after the first block of 20 trials, the operant level trials, would not be such a surprise to *S* (Matarazzo, Saslow, & Pareis, 1960). After the first 20 trials, each time *S* began a sentence with either "I" or "we" *E* said "good" at the end of the sentence for both reinforced groups, RC and RN.

For the nonreinforced groups UC (unreinforced control), and UN (unreinforced naming), *E* remained quiet throughout the 120 trials.

The color-naming task performed by the color-naming groups, RN and UN, consisted of naming colors during the 10-sec. intertrial period. The colors appeared in the memory drum every 1.43 sec. The non-color-naming groups, RC and UC, did not perform any task during the 10-sec. intertrial interval.

Following the presentation of the 120 cards all *S*s received the postconditioning interview, which was a series of questions read to *S* by *E* one at a time, with the answers to the questions being written by *S* on a blank sheet of paper. The interview used was patterned after that used by Dulany (1962).

Two graduate students in psychology were used as judges to rate the interviews obtained from *S*s. The judges rated the interviews of the *S*s for BI, RH, and BH. The criteria used were those given by Dulany (1962).

Results

The 120 trials were divided into six blocks of 20 trials each. The first block of trials was considered the operant level for each *S*. Number of "I-we" sentences was taken as the dependent variable. An analysis of variance of operant levels indicated that the color-naming variable was significant, $F\ (1, 96) = 7.62$, $p < .01$, with a higher operant level for the non-color-naming groups, but that the reinforcement variable, $F\ (1, 96) = 3.89$, $p > .05$ and the Reinforcement × Color-Naming interaction, $F\ (1, 96) = 0.00$, $p > .05$ were both nonsignificant.

An analysis of covariance was made using operant level as the control variable and involving the scores of all four groups on three variables: A, reinforcement; B, intertrial activity; and C, blocks of trials. The results of the analysis show that there was a significant overall reinforcement effect (A), $F\ (1, 95) = 12.57$, $p < .001$, i.e., the reinforced and nonreinforced *S*s differed significantly in the total number of "I-we" sentences produced. The intertrial activity variable (B) was not significant, $F\ (1, 95) = 0.06$, $p > .05$, nor was the interaction of intertrial activity and reinforcement (A × B), $F\ (1, 95) = 1.30$, $p > .05$.

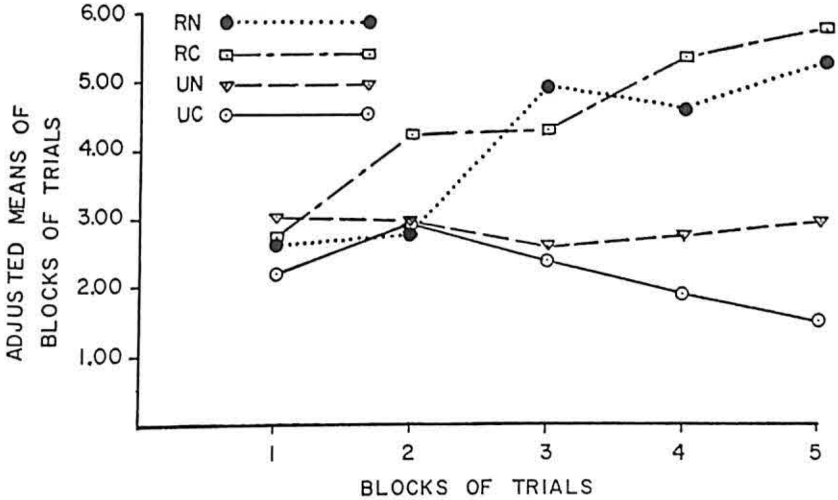


FIG. 1. Means for the four groups on the five blocks of trials adjusted for operant rates.

Both blocks of trials (C), F (4, 380) = 4.37, $p < .01$, and the interaction of blocks of trials with reinforcement (A × C), F (4, 380) = 8.59, $p < .001$, were significant. Neither the interaction of intertrial activity and blocks of trials (B × C), F (4, 380) = 1.08, $p > .05$, nor the three-way interaction of reinforcement, intertrial activity, and blocks of trials (A × B × C), F (4, 380) = 0.95, $p > .05$, was significant. Figure 1 presents the means for the four groups on the five blocks of trials adjusted for operant rates.

The degree of interjudge agreement for 100 interviews was 97% for BH, 97% for RH, and 85% for BI. The sum of the ratings for the two judges on each measure was used to correlate with conditioning score for S.

Table 1 presents correlations between a conditioning score obtained from the analysis of covariance, consisting of the individual's adjusted score on the last block of trials (adjusted for operant level), with levels of BH, RH, and BI for the two reinforced groups, RN and RC. The significant correlation between conditioning and RH in Group RC indicates that an increase in "awareness" was significantly related to an increase in conditioning in Group RC. (Smaller values for RH indicate greater degrees of "awareness.") The nonsignificant correlation between conditioning and level of RH in Group RN indicates that the degree of awareness was *not* significantly related to conditioning in the color-naming group. The difference between the correlations for Groups RN and RC is significant, $Z = 2.83$, $p < .01$. None of the other correlations were significant.

Chi square was used to compare the two reinforced groups, RC and RN, with respect to levels of RH. The value of chi square obtained (using a

TABLE 1

CORRELATION COEFFICIENTS OF LEVELS OF BH, RH, AND BI WITH CONDITIONING SCORES FOR GROUPS RN AND RC

Group	BH	RH	BI
RN	−.13	.06	−.06
RC	−.25	−.66**	−.30

** $p < .01$.

TABLE 2

RH Values (Summed Scores of Two Judges) for Groups RN and RC

Group	RH Level						
	2	3	4	5	6	7	8
RN	3	1	8		11		2
RC	2		7	2	14		

collapsed contingency table), X^2 (2) = .76, $p > .05$, indicated that Groups RC and RN did not differ with respect to the distribution of levels of RH. Table 2 presents the distribution of frequencies in the RH categories for the two reinforced groups.

Discussion

As indicated above, the Dulany-Spielberger position holds that the "awareness" on the part of *S* of the contingency between his verbal response and *E*'s reinforcer is a necessary condition for VOC. Dulany's (1962) postulates would predict significant correlations between BI, BH, RH, and conditioning. They would also predict that the magnitude of the correlations of these variables with conditioning scores would be of the order BI > BH > RH. From the Dulany (1962) theory, we would also predict that if the color-naming procedure interferes with the hypothesis-formulating behavior of *S* in any way, either through BI, BH, or RH, it should also interfere to a comparable degree with the conditioning score of *S*.

In checking these predictions from the Dulany-Spielberger interpretation against the results of the present experiment, we find that certain of them are not confirmed. In the first place, the color-naming task required between trials for Group RN did *not* adversely affect the conditioning of that group as would be expected from the theory. The analysis of covariance and Fig. 1 indicate that although the overall conditioning results show that there was learning in *both* reinforced groups (RN and RC), the nonsignificant A × B × C interaction indicates that the reinforced groups did not differ with respect to the amount of learning. This can be seen clearly from the graph in Fig. 1. However, the chi-square analysis showed that the distribution of RH ratings did not differ for Groups RN and RC. Thus, the two reinforced groups had the same distribution of RH ratings, but the correlation analysis (Table 2) indicates that the degree of "awareness" of the individual *S*s in Group RC as indicated by RH ratings was significantly correlated with the degree of conditioning, while there was no such relationship found for *S*s in Group RN. The degree of relationship between RH ratings and conditioning did differ significantly for Groups RN and RC. These findings suggest that the nature of the experimental task, together with the interview afterward, results in a certain distribution of degree of "awareness" among *S*s in a reinforced group, but that the degree and direction of the relationship between the level of "awareness" of individual *S*s and their degree of conditioning is dependent upon the degree to which *S*s are able to formulate hypotheses during the acquisition phase of the experiment. These findings would not be expected on the basis of the Dulany-Spielberger analysis of the VOC situation.

The further prediction from the Dulany theory that the magnitude of the correlation of the awareness variables with conditioning scores would be of the order BI > BH > RH was also not confirmed with our data. For Group RN the order was BH > BI > RH, and for Group RC the order was RH > BI > BH.

Although our results are contrary to what would be expected on the basis of the Dulany-Spielberger interpretation of VOC, they do seem to be consistent with what would be expected from the Verplanck-Krasner position. The Verplanck-Krasner position holds that there is no necessary relationship between degree of conditioning and the level of "awareness" of *S*s of a contingency between *S*'s response and *E*'s reinforcer. This interpretation holds that the action

of the reinforcer is direct, i.e., not mediated by cognitive processes, and that there may or may not be a relationship between the performance of *S* and his verbalization about his performance, i.e., his "awareness" of it, depending on the particular circumstances of the experiment. This is essentially the position of Skinner (1953, 1957). It would seem from our results that Verplanck (1962) and Krasner (1962) are correct in assuming that there is no necessary relationship between VOC scores and degree of awareness.

We would conclude, then, that although in the typical VOC experiment a positive correlation is found between level of "awareness" and amount of conditioning, this is due to the opportunity *S*s have during the acquisition phase of the experiment to formulate and test hypotheses regarding the relationship between *E*'s reinforcer and *S*'s behavior, but that there is no *necessary* relationship between "awareness" and such learning. When conditions are such that *S* is not able to formulate hypotheses during acquisition, the same degree of learning may be exhibited as when *S* does have the opportunity to formulate hypotheses, indicating a direct strengthening effect of the reinforcer, which is not correlated with the degree of *S*'s awareness. However, when such *S*s are later interviewed, as a result of the nature of the experimental task and the interview, a certain distribution of "awareness" levels will be exhibited by *S*s. But this "awareness" on the part of individual *S*s, which may indeed be produced by the interview itself, need not be related to the degree of conditioning exhibited by the particular *S*s during the experiment.

REFERENCES

BINDER, A., & SALOP, P. Reinforcement and personality factors in verbal conditioning. *J. Psychol.*, 1961, **52**, 379–402.

DULANY, D. E. Hypotheses and habits in verbal "operant conditioning." *J. abnorm. soc. Psychol.*, 1961, **63**, 251–263.

DULANY, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. *J. Pers.*, 1962, **30** (Supple.), 102–129.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

KRASNER, L. Reported awareness in verbal conditioning as a function of experimental conditions and subjects' personality. Paper read at Western Psychological Association, San Francisco, April 1962.

MATARAZZO, J. D., SASLOW, J., & PAREIS, E. N. Verbal conditioning of two response classes: Some methodological considerations. *J. abnorm. soc. Psychol.*, 1960, **61**, 190–206.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

SKINNER, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

SPIELBERGER, C. The role of awareness in verbal conditioning. *J. Pers.*, 1962, **30** (Supple.), 73–101.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

VERPLANCK, W. S. Unaware of where's awareness: Some verbal operants—notates, monents, and notants. *J. Pers.*, 1962, **30** (Supple.), 130–150.

(Received December 30, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 158–161

THE PARTIAL-REINFORCEMENT EFFECT SUSTAINED THROUGH BLOCKS OF CONTINUOUS REINFORCEMENT IN CLASSICAL EYELID CONDITIONING[1]

SALLY L. PERRY[2] AND JOHN W. MOORE

University of Massachusetts

The partial-reinforcement effect (PRE) was demonstrated in human eyelid conditioning by 2 groups of 24 Ss each: one group received 80 acquisition trials under 50% random reinforcement. The second group received 100% reinforcement. Percentage CRs over 20 extinction trials was highest for the first group. A third group was shifted from 50% to 100% reinforcement after Trial 40 and then extinguished after Trial 80. Covariance adjustments of the extinction performance of this group for acquisition performance and for intertrial blink rate indicated that the PRE was sustained through the block of continuous reinforcement. A modification of the Humphreys-expectancy version of the discrimination hypothesis could predict this result.

Discussion of the results of some recent experiments (e.g., Spence, Rutledge, & Talbott, 1963) involving extinction of the conditioned eyelid response of human *S*s has favored a Humphreys-expectancy version of the discrimination hypothesis (Lawrence & Festinger, 1962, p. 14) to account for the partial-reinforcement effect (PRE). Any application of the discrimination hypothesis must be questioned in light of experiments in instrumental conditioning, with both human and rat *S*s, which have demonstrated that the degree of shift in reinforcement schedule from acquisition to extinction has little effect upon the PRE (e.g., Capaldi & Capaldi, 1963; Jenkins, 1962), and, in the extreme ease, that the PRE is sustained through blocks of continuous reinforcement (Theios, 1962). The present experiment was designed to determine whether the prediction of the discrimination hypothesis that the PRE should *not* be sustained through blocks of continuous reinforcement is at least true for human eyelid conditioning.[3]

METHOD

The details of the apparatus have been described elsewhere (Mattson & Moore, 1964). Eighty-one student volunteers served as *S*s in three groups. Group 50-50 received 50% partial reinforcement for 80 acquisition trials. Group 50-100 received 50% partial reinforcement during the first 40 acquisition trials and was then shifted to 100% reinforcement for another 40 trials. Group 100-100 received 100% reinforcement for 80 acquisition trials. All groups received 20 extinction trials with no UCS presentations. Half of the *S*s in each of these groups were males and half were females. Nine *S*s (3 from each group) were discarded for failure to give at least three CRs in the last 10 acquisition trials. Of the remaining 72 *S*s, 12 were assigned to each of the six cells in a 3 × 2 randomized design in which schedule of reinforcement (50-50, 50-100, and 100-100) was orthogonal to *S*'s sex.

[1] This investigation was supported in part by Public Health Service Grant MH-7630-01 from the National Institute of Mental Health and was carried out by the first author in partial fulfillment of the requirements for the MS degree. We wish to thank L. E. Price for providing some of the apparatus.

[2] Now at Northeastern University.

[3] The arguments involved in the derivation of this prediction have been outlined by Theios (1962).

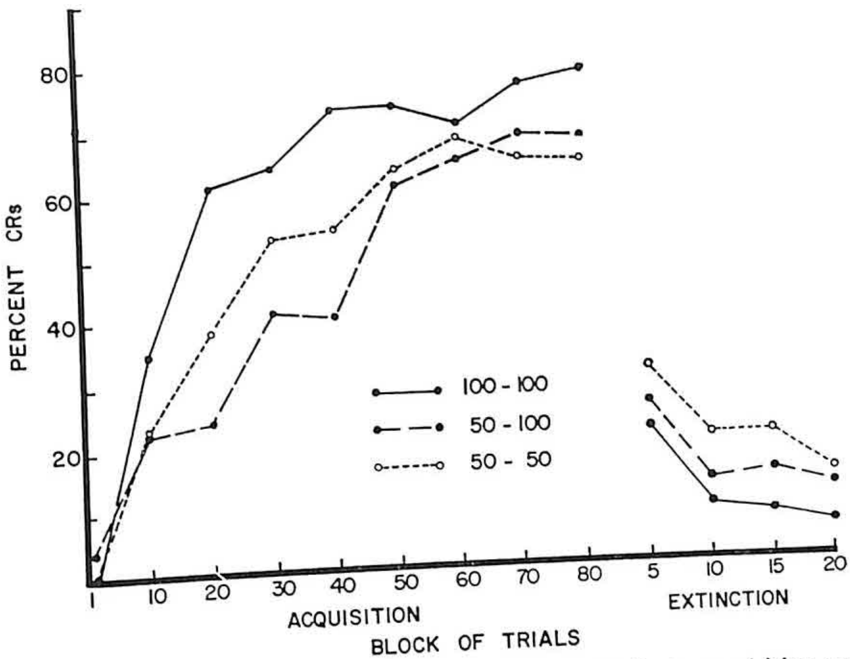


FIG. 1. Mean percentage CRs as a function of 10-trial blocks in acquisition and 5-trial blocks in extinction with schedule of reinforcement as the parameter. (The first point refers only to the first trial.)

The *S*s were given "neutral" instructions over an intercom immediately before the first trial. The CS was an 800-cps tone of 70 db. SPL and of 550 msec. duration. The UCS was a 1.0 lb/sq in air puff of 50 msec. duration. On reinforced trials the CS-UCS interval was .5 sec. The 50% reinforcement schedule was randomized so that (*a*) no more than 4 reinforced or nonreinforced trials occurred consecutively and (*b*) each block of 10 trials contained 5 reinforced and 5 nonreinforced trials. The intertrial intervals were 15, 22.5, and 30 sec. randomly distributed over the 100 trials of the experiment.

A CR was defined as any deflection of the recording pen of at least 1 mm. from the base line and occurring in the range of 150 to 500 msec. after CS onset. Intertrial responses were recorded in order to increase the precision of statistical analysis of conditioning scores (Mattson & Moore, 1964). The criterion for an intertrial response was any deflection and return of the recording pen of at least 1 mm. The blink rate for each intertrial interval was obtained by summing the number of blinks in the interval and dividing by the number of successive 500-msec. segments within that interval. If no more than one response is counted in a 500-msec. segment, this procedure yields a percentage score which can be placed on the same scale of measurement as percentage CRs.

Results and Discussion

Figure 1 presents the percentage of CRs for the three groups as a function of trial block. Groups 50-50, 50-100, and 100-100 gave 64, 68, and 78% CRs, respectively, on the last block of acquisition trials, but the order of responding in extinction for the three groups, from highest to lowest, was a complete inversion of this order with Group 50-100 approximately midway between Groups 50-50 and 100-100.[4]

[4] Results of experiments by Gormezano and Moore (1962) and Moore and Gormezano (1963), which are based on procedures comparable to those in the present study, suggest that most "voluntary" responders would be among those *S*s who gave more than 80% CRs in acquisition. The number of *S*s in Groups 50-50, 50-100, and 100-100 who gave more than 80% CRs in acquisition was one, two, and seven, respectively.

For all three groups, intertrial blink rate (not shown) remained essentially constant at a value of 20% in acquisition and 15% in extinction.

For statistical analyses an arcsine transformation was applied to each of the following percentage scores for each *S:* (*a*) CRs based on 80 acquisition trials, (*b*) CRs based on the last 40 acquisition trials, (*c*) CRs in extinction, and (*d*) intertrial blink rate in extinction. For ease of presentation, these five measures are henceforth referred to as (*a*) acquisition score, (*b*) asymptotic acquisition score (*c*) extinction score, and (*d*) extinction blink rate, respectively.

An analysis of variance of extinction scores yielded a significant main effect due to the schedule of reinforcement in acquisition, $F(2,66) = 3.91$, $p < .05$. This main effect was even more significant, $F(2, 65) = 11.37$, $p < .001$ when extinction scores were adjusted for acquisition scores by an analysis of covariance. With an additional covariate, extinction blink rate, schedule of reinforcement was still highly significant, $F(2,63) = 12.97$, $p < .001$. Since the asymptotic acquisition scores might be considered the more appropriate covariate for extinction, this measure and extinction blink rate were used in an analysis of covariance of extinction scores which again indicated a significant effect due to schedule of reinforcement, $F(2, 63) = 11.26$, $p < .001$. The correlation between extinction scores and extinction blink rate, averaged over the six cells of the design, was only .28.

Using acquisition scores and extinction blink rate as covariates, the adjusted mean extinction scores for the three main groups were 42.19 (50-50), 37.35 (50-100), and 20.04 (100-100). The difference between the adjusted means of Groups 50-50 and 50-100 was not significant, but Groups 50-100 and 100-100 differed significantly from each other, $F(1, 63) = 15.84$, $p < .001$. Using the asymptotic acquisition scores and blink rate in extinction as the covariates, the adjusted mean extinction scores were 37.95 (50-50), 34.95 (50-100), and 26.99 (100-100). The mean of Group 50-100 was again closer to that of Group 50-50 than to that of Group 100-100, but the difference between 50-100 and 100-100 was not significant.

These analyses suggest that the PRE was sustained through the block of continuous reinforcement, and, in light of the analyses of covariance, any attenuation of the PRE by continuous reinforcement was slight. If one assumes that resistance to extinction depends not only on *S*'s awareness of the change in reinforcement schedule from acquisition to extinction but also upon his *interpretation* of the change, then an expanded form of the Humphreys-expectancy type discrimination hypothesis might account for the results. The *S*s in Group 50-100 may have interpreted the onset of extinction as a reinstatement of partial reinforcement rather than as the start of extinction, and the extent to which such an interpretation was made should have increased the level of responding of these *S*s relative to those in Group 100-100. The *S*s in the latter group, having had no previous experience with partial reinforcement, could not be expected to have interpreted the shift as the introduction of a partial-reinforcement schedule.

REFERENCES

Capaldi, A. B., & Capaldi, E. J. Influence of patterns of events in sequence on resistence to extinction. *Psychol. Rep.*, 1963, 12, 343–347.

Gormezano, I., & Moore, J. W. Effects of instructional set and UCS intensity on the latency, percentage, and form of the eyelid

response. *J. exp. Psychol.*, 1962, **63**, 487–494.

Jenkins, H. M. Resistence to extinction when partial reinforcement is followed by regular reinforcement. *J. exp. Psychol.*, 1962, **64**, 441–450.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcement.* Stanford: Stanford Univer. Press, 1962.

Mattson, M., & Moore, J. W. Intertrial responding and CS intensity in classical eyelid conditioning. *J. exp. Psychol.*, 1964, **68**, 396–401.

Moore, J. W., & Gormezano, I. Effects of omitted versus delayed UCS on classical eyelid conditioning under partial reinforcement. *J. exp. Psychol.*, 1963, **65**, 248–257.

Spence, K. W., Rutledge, E. F., & Talbott, J. H. Effect of number of acquisition trials and the presence or absence of the UCS on extinction of the eyelid CR. *J. exp. Psychol.*, 1963, **66**, 286–291.

Theios, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 1–6.

(Received December 30, 1963)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 162–169

BASES FOR PREFERENCES AMONG THREE-OUTCOME BETS [1]

SARAH LICHTENSTEIN [2]

University of Michigan

This experiment compared the moment functions of a bet as predictors of choices among bets with the subjectively expected utility (SEU) maximization model. Bet parameters of Expected Value (EV), Variance (V), Skewness (Sk), and probabilities were independently varied in 182 3-outcome bets. Each of 12 Ss bid on each bet and played some bets for real money. Ss preferred high EV and low V, but had no preferences within Sk or probability levels. Sets of bets existed which were equal in all parameters; within such sets, Ss preferred bets with the largest least likely amount. The moment-function approach was rejected; the SEU model was not. A lexicographic ordering of variables was suggested.

Research on decision making in gambling settings has focused in recent years on a model called the subjectively expected utility (SEU) maximization model. This model states that choices among bets can be predicted from the maximization of the SEU function, SEU $= \psi_1 u_1 + \psi_2 u_2 + \cdots + \psi_n u_n$, where there are n possible outcomes of a bet, the first outcome has utility u_1 and subjective probability ψ_1, and so on. Research on the model has concentrated on scaling the utility and subjective probability curves for individuals or groups of *S*s. For reviews of this literature see Edwards (1954a, 1961).

Criticism of the SEU model has come from Coombs and Pruitt (1960), who feel that the SEU model is undesirable because of its relative invulnerability to experimental refutation, in that utility and subjective probability curves may be found to fit almost any set of data. Accordingly, they argue against SEU model not by collecting evidence which refutes it, but by proposing an alternative set of variables which might be used as the basis of a new model, and which they believe is more vulnerable to experimental evidence.

Coombs and Pruitt point out that the value or Expected Value (EV) of a bet is the first of a series of moment functions which uniquely describe the bet, when the bet is viewed as a probability-density distribution. Other moment functions are Variance (V), Skewness (Sk), and Kurtosis (K). EV has long been recognized as an important determinant of choices among bets. Many authors (e.g., Allais, 1953; Edwards, 1954b; Royden, Suppes, & Walsh, 1959) have proposed that V is also important. For two-outcome bets, Sk is a simple function of the probabilities, so that probability preferences might be relabeled Sk preferences. Coombs and Pruitt proposed the moment functions as a possible set of parameters useful in predicting choices among bets.

[1] This paper is based on a PhD dissertation submitted to the Department of Psychology of the University of Michigan. The author is greatly indebted to Ward Edwards for his encouragement and guidance throughout the investigation. This research was supported in part by the Air Force Office of Scientific Research, Grant No. AFOSR 192-63.

[2] Now at Southern Illinois University, Carbondale, Illinois.

The SEU model states that a function of two subjective distributions, utility and subjective probability, determines choices among bets. In contrast, Coombs and Pruitt suggest that the entire shape of the objective distribution determines choices. The moment functions are a set of parameters summarizing that shape. Either of these general models can encompass so much data that experimental refutation is difficult. But both these approaches are vulnerable to experimental refutation when details of the model are specified, as for example Davidson, Suppes, and Siegel (1957) have done for the SEU model, or when a small set of moment functions, EV, V, and Sk, limits the objective-distribution approach.

A decision between the moment functions and the SEU model, however, is hampered by a lack of systematic, empirical information about the effects of a number of bet variables upon *Ss'* choices. This experiment was designed to supply this lack by exploring the relevance and relative importance of the following bet parameters: the first four moment functions, EV, V, Sk, and K, as defined in Table 1, and the set of probabilities of a bet.

TABLE 1

FORMULAE FOR MOMENT FUNCTIONS

	Moment Notation	Definitional Formula
Expected Value	μ_1	$\sum_i p_i \$_i$
Variance	$\mu_{2,c}$	$\sum_i p_i(\$_i - \mu_1)^2$
Skewness	$\frac{\mu_{3,c}}{[\mu_{2,c}]^{\frac{3}{2}}}$	$\frac{\sum_i p_i(\$_i - \mu_1)^3}{[\mu_{2,c}]^{\frac{3}{2}}}$
Kurtosis	$\frac{\mu_{4,c}}{[\mu_{2,c}]^2}$	$\frac{\sum_i p_i(\$_i - \mu_1)^4}{[\mu_{2,c}]^2}$

Note.—p_i is the probability of the ith outcome; $\$_i$ is the amount of money to be won or lost.

In order to evaluate properly the effectiveness of these bet parameters, it is important that they be mutually independent in the experimental design. This has not been true in previous research. Edwards (1953), in exhibiting probability preferences, confounded probability, V and Sk. Coombs and Pruitt (1960) confounded probability with Sk. Such confounding is inevitable when bets having only two outcomes are used, since the equation for Sk then reduces to a simple function of the probabilities

$$\left(\text{Sk} = \frac{1 - 2p_1}{\sqrt{p_1 p_2}}\right).$$

This study used three-outcome bets, permitting independent variation of probabilities, EV, V, and Sk, with K confounded. The amounts of money, $\$_i$, needed to define a bet, given the values p_i, EV, V, and Sk, may be found by solving the equations in Table 1 by approximation techniques.[3] In some cases there are multiple solutions of $\$_i$, so that several different bets exist which are equal in their probabilities, EV, V, and Sk, but which differ in wins and losses. For example, the bet, "1/8 chance to win \$1.82; 2/8 chance to lose \$2.49; 5/8 chance to win \$.63," and the bet, "1/8 chance to lose \$3.00; 2/8 chance to lose \$1.23; 5/8 chance to win \$1.08" both have EV = 0, V = 2.25, and Sk = −.9. When probabilities in units of eighths are used, no bets exist when Sk is outside the range ±2.26. Within this range, from zero to four bets exist for each choice of probabilities, EV, V, and Sk.

[3] The author is indebted to L. J. Savage for suggesting the graphic method of solution used to obtain the bets.

The existence of several bets all having the same probabilities, EV, V, and Sk is a powerful feature of the experimental design. It permits a direct assessment of the value of the moment functions as a useful set of predictive variables. If the variation of *S*'s responses within such a set is as large as the variation between such sets, the moment functions will be judged inadequate, unless K, the next such function, is related to the variations in response.

To measure *S*s' preferences, a single stimulus, real gambling bidding method was employed, both because the more usual paired-comparison techniques are impractical when large numbers of bets are used, and as a means of increasing *S*s' motivation. Previous research with this bidding method (Slovic, Lichtenstein, & Edwards, in press) indicated that it provides powerful ratio scale data which, when reduced to order relationships, are similar to data obtained by paired comparisons. Its greater economy over paired-comparison methods and its attractiveness to *S*s further argue for its use.

Method

Stimulus Material

The stimuli used were 182 three-outcome bets and 7 two-outcome bets. Each bet was presented five times. The three-outcome bets were computed from combinations of the following levels of the bet parameters. The number of different bets at each level is shown in parentheses.

Probability.—All five possible triples based on the unit of an eighth were used: 1/8, 1/8, 6/8 (16 bets); 1/8, 2/8, 5/8 (96); 1/8, 3/8, 4/8 (28); 2/8, 2/8, 4/8 (18); and 2/8, 3/8, 3/8 (24).

EV.—Three levels: −83¢ (39 bets), zero (104), and +83¢ (39).

V.—Four levels: .25 (107 bets), 1.00 (25), 2.25 (25), and 4.00 (25).

Sk.—Five levels: −1.8 (16 bets), −.9 (42), zero (66), +.9 (42), and +1.8 (16).

K.—The bets also varied in K, ranging from 1.3 to 5.2. K remains constant with changes in EV and V, and with changes in the sign of Sk.

Number of bets.—When probability and Sk are varied, while EV and V are held constant, 41 different bets may be calculated. A linear transformation on any bet produces a new bet with the same probabilities and Sk, but new EV and V. Thus when EV and V are varied, a total design of 492 bets (41 × 3 EV × 4 V) is produced. The 182 three-outcome bets used in the experiment comprise a subset of the possible 492 bets. This subset was chosen so that only second- and third-order interaction effects would be lost to the analysis. The bets used are listed in Lichtenstein (1962).

The seven two-outcome bets had EV = 0 and V = .25, with Sk values ranging from −2.26 to +2.26.

The design of this study is incomplete, due both to the variable number of bets which can be computed for given values of the parameters, and to the reduction in the number of bets used from 492 to 182.

Response Method

The bets were presented one at a time to each *S*. The bets were shown on 4 × 6 in. cards as pie diagrams, with the amount of money corresponding to each outcome written inside the sectors of the pie, e.g., "win $1.00," "lose $.50," or "win nothing."

For each bet, *S* was asked to state the largest amount of money he would be willing to pay *E* in order to play the bet. If *S* did not like the bet well enough to pay for it, he was asked to state the smallest amount of money *E* had to pay him before *S* would play the bet. The *S* was told that this amount of money, his bid for the bet, would be compared with bids made by other *S*s in former experiments, and that his bid would be accepted and the bet played only if it fell in the upper quarter of previous bids for the bet. This competitive bidding was a fiction used to motivate *S*s to give careful responses; there was no such comparison group. The *E* actually accepted a random third of *S*'s bids, so that *S* did not receive any consistent feedback useful to him in developing a strategy of play.

When a bid was declared acceptable by *E*, the bid amount was exchanged and the bet was played. When the bid was refused, the bet was not played.

The bets were played by spinning a balanced aluminum pointer placed over a plastic disc 8 in. in diameter. Five discs cor-

responding to the five probability combinations for three-outcome bets, and four discs with two-outcome probabilities, were prepared by painting the backs in colors red, blue, and green, in pie-slice sectors.

Procedure

Subjects.—The *S*s were 12 male volunteers, undergraduates from the University of Michigan.

Sessions.—The *S*s were run singly in seven 2-hr. sessions spread over a 2-wk. period. Five of the sessions were devoted to the experimental bets; in each session *S*s bid on 189 bets, a single complete replication. The bets were presented in a different random order for each session; all *S*s saw the bets in the same sequence.

Two sessions, the fourth and last, were adjustment sessions to control the total amount of each *S*'s winnings. In these sessions *S* made paired-comparison choices between pairs of bets. The *S* was told he would play a random half of his choices for money. Actually *S*s who had already won more than the budgeted amount played mostly negative EV bets, while *S*s who had lost money or had won less than the budgeted amount played mostly positive EV bets.

Financial exchanges.—All *S*s were told at the beginning of the experiment that they would be gambling for real money, and that while the game was biased in their favor, there was a possibility that they could lose some of their own money. Financial exchanges during a session were made with poker chips. The *S*s were advanced $30.00 in chips at the beginning of each session and required to return the same amount at the end of the session.

Results

Data analysis.—Analysis of the data was based on the amount of money each *S* bid for each replication of each bet. A positive number means *S* offered to pay *E* for the bet; a negative number means *S* required *E* to pay *S* before he would play the bet. Because of the incompleteness of design, when means are computed across all the bets used in the study certain levels of the parameters are overrepresented. In order to avoid drawing false conclusions from data in which several levels of the variables are unevenly represented, five small subsets of the data, each factorially complete, and each representing the variation of two parameters while two are held constant, were subjected to analyses of variance. Thus any conclusion drawn from the total data was checked in two or three narrower but purer settings.

Subject consistency.—The *S*s bid carefully and with moderate consistency. The average standard deviation of the five bids a single *S* made to a single bet in five replications was 14¢, for all *S*s and all bets. The most consistent *S* had a per-bet average standard deviation of 5¢; for the most variable *S* this figure was 23¢.

All *S*s made clear by their comments during the experiment that they accepted the competitive bidding fiction and believed in the game; they were trying very hard to play the game wisely and to win money. When *S*s were winning, they were happy, cheerful, and "chatty." When *S*s were losing they became silent, moody, and often rude, picking arguments over trivia.

Expected Value.—All *S*s were sensitive to the EV of the bets. The first column in Table 2 shows the mean amount bid for each EV level. Over the entire experiment, the average bid was 12¢ below EV. Only one *S* often bid too high; 68% of his bids averaged over replications were above EV. For the other 11 *S*s this figure was 16%.

Variance.—The *S*s were also sensitive to changes in V, as shown in Table 2. The *S*s bid more for low V bets than for high V bets. The data were then divided into 36 subsets comprising each of the 12 *S*s and each of the 3 EV levels. The bets within each subset were constant in EV but differed in V, Sk, and probabilities. Each subset was examined for V

TABLE 2

MEAN AMOUNT BID BY ALL *S*s FOR EACH LEVEL OF EACH PARAMETER

Expected Value		Variance		Skewness		Probabilities	
Level	Mean Bid	Level	Mean Bid	Level	Mean Bid	Level	Mean Bid
−83¢	−91¢	.25	−07¢	−1.8	−06¢	1/8, 1/8, 6/8	−09¢
0	−11¢	1.00	−12¢	−.9	−09¢	1/8, 2/8, 5/8	−13¢
83¢	67¢	2.25	−17¢	0	−13¢	1/8, 3/8, 4/8	−09¢
		4.00	−25¢	.9	−12¢	2/8, 2/8, 4/8	−08¢
				1.8	−12¢	2/8, 3/8, 3/8	−10¢

Note.—A negative bid means *S* required *E* to pay *S* before *S* would play the bet.

preferences. Of these subsets, 21 sets showed a simple ordering from low V to high (10 sets with zero EV, 6 with positive EV, and 5 with negative EV). Of the other 15 sets, 11 had no clear preference order but differed from the order shown in Table 2 by only a few cents, while only four sets had clearly different orders.

Skewness.—Skewness had no noticeable effect on *S*s' bidding. Inspection of the data revealed that no single *S* had a clear preference ordering among the five Sk levels, nor was there any small effect cumulative across *S*s. Table 2 shows the mean amount bid at different Sk levels.

Probability.—There was no evidence for probability preferences for any single *S*, nor for all *S*s cumulatively, as shown in Table 2.

Analyses of variance.—The five subsets of data subjected to analysis of variance (EV × V, EV × Sk, EV × Probability, V × Sk, and V × Probability) confirmed the above results. EV and V main effects were extremely large, while Sk and probability showed no effects. The Subjects main effects were of moderate size. All interactions were insignificant.

Kurtosis.—Correlations between K and amount bid averaged over replications were computed for each *S* and for each EV level separately. These correlations ranged from −.32 to +.30, with a mean of +.02. More detailed analyses checking for possible curvilinear relationships, and separate analyses for each V level, indicated no relationship between K and amount bid.

Bets with equal independent variables.—The existence of multiple bets with constant probability, EV, V, and Sk provides an opportunity to test for bet preferences unrelated to the moment functions. These sets of bets were examined to see if differences in amount bid among the bets were large or interpretable. The design includes 12 such sets of four bets each, 6 sets of three bets each, and 46 sets of two bets each.

Consistent ordering of these bets within sets was shown; *S*s consistently preferred bets in which the amount of money associated with the least likely outcome was large and positive. For example, *S*s bid 20¢ more, on the average, for the bet, "1/8 to win \$1.83; 2/8 to lose \$2.49; 5/8 to win \$.63," than for the bet, "1/8 to lose \$3.00; 2/8 to lose \$1.23; 5/8 to win \$1.08." Both these bets have EV = 0, V = 2.25, and Sk = .9.

Table 3 summarizes these differences among bets of equal independent variables, under the hypothesis that *S*s will prefer bets in which the least likely outcome has the largest win. For sets of three and four bets, the

differences shown in Table 3 were taken between the bet with the largest win and the bet with the largest loss in the least likely outcome. The intermediate bets are not included in this summary, but were highly consistent with the reported results.

Among the 46 pairs of bets with equal independent variables, 18 sharply differ in least likely amount. Table 3 shows the 13¢ mean difference in amount bid between these bets. For the 28 pairs which are equal or nearly equal in least likely amount, the prediction was made that *S*s would base their preferences on the amount of the next least likely outcome (i.e., the outcome associated with the middle-sized probability). Table 3 shows that the mean difference in amount bid was small (4¢) for this group.

These consistent preferences among bets of equal EV, V, Sk, and probability were not related to the K values of the bets.

Effects of real gambling.—Detailed analyses of individual *S*'s data showed that *S*s' monetary success or failure had little or no effects on *S*s' bids. Gambling variables which failed to correlate with amount bid were: the total amount of money *S* had previously won or lost when he made a given bid, the amount of money won or lost within a single session, the amount of money won or lost on the last three gambles, and the number of successive bids refused by *E*.

Two-outcome bets.—Data from the seven two-outcome bets were examined for evidence of preferences among the bets, but little preference was found. Only six *S*s showed a difference in amount bid of 15¢ or more between the most and least preferred bets. These six *S*s had no common pattern in their orderings of the seven bets.

TABLE 3

DIFFERENCES IN MEAN AMOUNT BID WITHIN SETS OF BETS WITH EQUAL PARAMETERS

	No. of Sets	No. with Difference in Predicted Direction	(%)	Mean Difference
Quadruples of bets	12	12	(100%)	18¢
Triples of bets	6	6	(100%)	22¢
Pairs of bets differing in least likely amount	18	15	(83%)	13¢
Pairs of bets same in least likely amount	28	21	(75%)	4¢

Subject differences.—Finally, a global view of each *S*'s data for the entire experiment was taken to see if there were any striking differences among *S*s. Two *S*s were notable. Subject 2 often bid above EV for bets, and thus lost money in the bidding sessions. His unprofitable play, however, was not highly erratic; his mean standard deviation across replications was below the group average. He was the only *S* who preferred high V bets. Subject 6 carefully attended to the EV of bets and fairly consistently bid 3¢ below EV. His data are striking for their evenness, showing no effects attributable to any parameter other than EV. The other 10 *S*s, although varying slightly in different aspects of their performance, seem well represented by the group data.

DISCUSSION

Moment functions.—This study was based on the suggestion that the set of moment functions, EV, V, Sk, and K, might provide the basis for a decision-making model, as an alternative to the SEU model. The results must be evaluated in the light of this suggestion.

First, there was a sizable effect attributable to EV, as expected. Indeed, it is so certain that *S*s attend to EV, that failure to find a substantial EV effect would invalidate the study. Variance also had an effect on *S*s' behavior. Ten of the 12 *S*s showed a preference for low V bets. In a sense, this result replicates the results of Coombs and Pruitt (1960), who also found V preferences. But while Coombs and Pruitt's *S*s showed wide variation in their most preferred V level, most *S*s in the present study preferred low V bets. This difference may be due to the difference in motivating conditions in the two studies. In the Coombs and Pruitt study, *S*s were paid a salary and did not play any bets during the experiment. In contrast, *S*s in this study played one-third of the bets for real money, and believed that their responses were the major determinant of how much money they would win or lose. The fears of losing money under such conditions may account for the avoidance of the riskier, high V bets.

In the present study, neither the third moment, Sk, nor the fourth moment, K was related to *S*s' behavior. An even more serious obstacle to the moment-function approach was the finding of large and consistent differences in bidding among bets whose first three moment functions are equal. These findings lead to the rejection of the moment functions as a set of variables sufficient to build a model.

SEU model.—None of the three primary results, the effect of V, the effect of the least likely outcome, and the lack of probability preferences, contradict the SEU model. The preference for low V bets indicates that the utility curve for money is not symmetric in its extreme ranges; that is, that large losses appear larger than large wins. The effect of the least likely outcome indicates that *S*s overestimate small probabilities, so that the subjective probability curve is not linear with objective probability. A nonlinear subjective probability function applied to two-outcome bets suggests the existence of probability preferences, while the complex, nonmonotonic nature of three-outcome bets obscures probability preferences.

Probability preferences.—In this study, *S*s showed no probability preferences. Edwards (1953), using two-outcome bets, found probability preferences, but his bets confounded probability with V. Coombs and Pruitt (1960) also found probability preferences but they were weak and unsystematic relative to V preferences.

Least likely amount.—The unexpected finding of consistent preferences among bets of equal EV, V, Sk, and probabilities is of interest. Apparently *S*s lay undue emphasis on the amount of money associated with the smallest probability, bidding more for a bet when this amount is positive than when it is negative. This effect of the least likely amount is quite distinct from the effect of V. To illustrate this distinction, consider the following bets:

A. 1/8 to win $.12; 2/8 to win $.83; 5/8 to lose $.35.

B. 1/8 to win $1.00; 2/8 to win $.41; 5/8 to lose $.36.

These two bets have equal EV, V, and Sk. Within this pair, *S*s bid more for Bet B, the bet with the largest least likely amount. In contrast are the following bets:

A. 1/8 to win $.12; 2/8 to win $.83; 5/8 to lose $.35.

C. 1/8 to win $.50; 2/8 to win $3.50; 5/8 to lose $1.42.

These two bets have equal EV and Sk but differ in V. Here, *S*s bid more for Bet A, the bet with the smaller V, even though Bet C has the larger least likely amount.

These findings suggest that *S*s approach bets with a lexicographic ordering of relevant variables: EV is of primary importance. When deciding between bets of equal EV, *S*s prefer low V. When both EV and V are constant, *S*s attend to the least likely amount, avoiding a small chance of large loss.

REFERENCES

ALLAIS, M. Le comportement de l'homme rationnel devant le risque. Critique des postulats et axioms de l'ecole americaine. *Econometrica*, 1953, **21**, 503–546.

COOMBS, C. G., & PRUITT, D. C. Components of risk in decision making. *J. exp. Psychol.*, 1960, **60**, 265–277.

DAVIDSON, D., SUPPES, P., & SIEGEL, S. *Decision-making: An experimental approach.* Stanford, Calif.: Stanford Univer. Press, 1957.

EDWARDS, W. Probability-preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

EDWARDS, W. The theory of decision-making. *Psychol. Bull.*, 1954, **51**, 380–417. (a)

EDWARDS, W. Variance preferences in gambling. *Amer. J. Psychol.*, 1954, **67**, 441–452. (b)

EDWARDS, W. Behavioral decision theory. *Annu. Rev. Psychol.*, 1961, **12**, 473–498.

LICHTENSTEIN, S. *Bases for preferences among three-outcome bets.* (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1962, No. 63–391.

ROYDEN, H. L., SUPPES, P., & WALSH, K. A model for the experimental measurement of the utility of gambling. *Behav. Sci.*, 1959, **4**, 11–18.

SLOVIC, P., LICHTENSTEIN, S., & EDWARDS, W. Boredom-induced changes in preferences among bets. *Amer. J. Psychol.*, in press.

(Received January 2, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 170–179

APPARENT SPATIAL POSITION AND THE PERCEPTION OF LIGHTNESS [1]

JACOB BECK
Harvard University

2 studies investigated the relation of lightness perception to the perception of spatial position. The results confirm earlier findings that lightness perception may be affected by how an O perceives the surface to be oriented with respect to the illumination. The results fail to support the hypothesis, however, that the apparent position of a surface relative to the illumination is used as a basis for computing the albedo of a surface. Rather, the general hypothesis the studies appear to support is that processes of perceptual organization come into play as a result of the cue properties of stimuli which affect whether a variation in luminance will be seen as a difference in the illumination of the surface or as a difference in the lightness of the surface. Thus, an area of reduced surface luminance seen in one position as a shadow is, in another, seen as a gray surface color, in each case consistent with the apparent position of the surface.

Studies have shown surface lightness to change with changes in the apparent depth (Woodworth & Schlosberg, 1954, p. 444), slant (Katona, 1935; Mach, 1959, p. 209), or tilt (Hochberg & Beck, 1954) of a surface. Evidently, a description of lightness perception must take into account the apparent spatial position of a surface as well as the light reflected by the surface. The effects of apparent spatial position imply that not only the luminance relationships, but the cue properties of the stimulus influence lightness perception. The concern of the present experiments is how cues for spatial position affect the perception of lightness.

One hypothesis is to suppose that the apparent position of a surface enters into computing the albedo of a surface. The perception of lightness is assumed to be the result of an unconscious computing in which the intensity of the reflected light is evaluated with respect to the intensity of the registered illumination (Woodworth & Schlosberg, 1954, pp. 430–431). Change in the apparent position of a surface affects perceived lightness because a surface in the new position would have to have a lower (or higher) albedo to reflect the same amount of light to the eyes. Beck (1961, 1962), however, has argued that perceived lightness and illumination are not mutually related in the manner implied by the albedo hypothesis. Moreover, lightness perception of uniformly illuminated surfaces has been described without recourse to the impression of illumination as a cue (Helson, 1943; Stevens, 1961; Wallach, 1948).

A second hypothesis is to attribute the change in lightness to processes of perceptual organization that come into play in dealing with the distributional properties of reflected light. What is supposed is that operating together with sensory processes, such as contrast and adaptation, there are

[1] This research was supported in part by funds granted by Public Health Service Research Grant No. M-05120-02 to Harvard University, Center for Cognitive Studies and in part by National Science Foundation Grant GB-94 to Harvard University, Psychological Laboratories.

organizational processes which influence lightness perception by the way a stimulus pattern[2] is organized into a systematic whole. That is, as a result of the cues provided by the distribution and character of the luminance variations, the apparent direction of illumination and the apparent position of a surface, luminance variations may be perceived either as differences in illumination or as differences in lightness. For a coplanar surface and surround perceived to be illuminated uniformly, the hypothesis asserts that luminance variations would be perceived as differences in surface lightness enabling lightness perception to be expressed as a function of the distribution of luminances in the field (Stevens, 1961; Wallach, 1948). For a surface perceived to be illuminated directionally, however, the organization of the stimulus pattern may be supposed to operate to maximize the congruences between different cues. Thus, an area of reduced surface luminance seen in one position as a shadow, is, in another, seen as a gray surface color, in each case consistent with the apparent position of the surface.

The two hypotheses make different predictions concerning the effects of spatial position on perceived lightness. The albedo hypothesis asserts that perceived lightness is given by the equation $A = S/M$ (Woodworth & Schlosberg, 1954, p. 431) where M = the registered illumination, S = intensity of the reflected light, and A = the perceived lightness or albedo. This formula implies: (*a*) altering the apparent position of a surface so as to modify its apparent illumination is a sufficient condition for a change in surface lightness, and (*b*) for a given retinal stimulus there is a proportional variation between changes of apparent illumination and lightness. The second hypothesis asserts that lightness perception can be analyzed as a two-step process. First, sensory processes act to determine a stimulus pattern in accordance with the distribution of luminances reflected, and, second, as a result of the cue properties of the stimulus, the stimulus pattern is organized in terms either of differences in surface lightness or differences in illumination. Implications of this hypothesis are: (*a*) perceived lightness is affected only if a change in the apparent position of a surface induces a new organization of the stimulus pattern, and not simply because of the altered impression of illumination, and (*b*) the range of lightness perception is a function of the stimulus pattern determined by the distribution of luminances reflected and not of the amount of change in the apparent illumination.

The experiments to be reported provide a test of the albedo hypothesis and its implications. They also serve to clarify the factual background of the relation between the perception of lightness and of spatial position.

Study I

Study I was designed to examine the stimulus conditions controlling lightness changes in Mach's (1959, p. 209) demonstration.

Method

The experiments were conducted in a darkroom. A folded card was placed on a table behind a cardboard partition so that

[2] Since the hypothesized organizational process may take place only after a number of prior perceptual functions, such as contrast processes, contour processes, and so forth, the phrase stimulus pattern is used as a neutral term to avoid implying that the organizational process operates directly on either the array of light stimulating the eye or the proximal stimulus pattern, vid. the discussion section of this report.

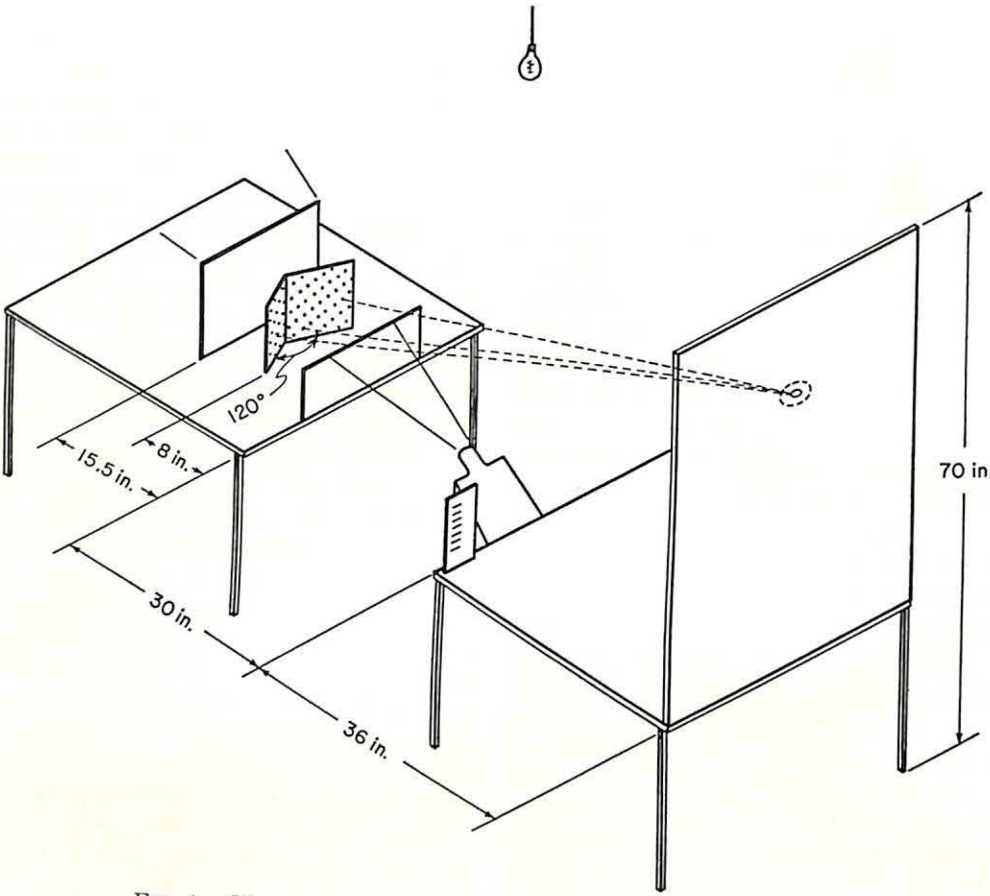


FIG. 1. Illustration of the experimental arrangement in Study I.

left side was illuminated while its right side was shadowed. The table was covered with a black cloth. Apertures cut in the screen fixed to one end of the table allowed *O*s to see the target with either monocular or binocular vision. Changes in the apparent position of the target were induced by drawing rows of dots making an angle of 13° with the horizontal on each side of the target. When cues for slant were eliminated by limiting *O* to monocular vision, the presumption to see the dots arranged in straight lines caused the target to appear to stand flat directly behind the partition. When the target was viewed binocularly, the target was seen in its correct spatial position and the rows of dots were seen as slanted.

Four experiments were conducted. In Exp. I and II, the illumination (150 w.) came from a ceiling lamp. The illuminance of the illuminated side of the target was 4.2 footcandles (ftc.) and of the shaded side of the target 2.8 ftc. When the target was seen as folded, cues indicated that the left side was directly illuminated and the right side was indirectly illuminated. When the target was seen as flat, both sides appeared to be illuminated equally. The illuminance of the left and right sides of a white paper of the same reflectance as the target and subtending the same visual angle when placed flat directly behind the partition was equal to 4.2 ftc. Thus, changes in the apparent spatial position of the target altered the apparent illumination of the shaded side of the target but not of the illuminated side. In Exp. III and IV, light from a projector located below *O*'s line of sight was added to the overhead light. The projector beam actually missed the target falling only on the cardboard partition and the wall behind. However, the target when apparently flat appeared to receive the bright light of the projector. The illuminance of the right and

TABLE 1

MEDIANS OF *O*s' LIGHTNESS MATCHES OF THE SHADED AND ILLUMINATED SIDE OF THE MACH CARD WITH DIFFERENT APPARENT POSITIONS

Exp.	Shaded Side		Exp.	Illuminated Side	
	Apparently Flat	Apparently Folded		Apparently Flat	Apparently Folded
I	8.5	9	II	9	9.25
III	6.75	8	IV	8.5	8.5

Note.—A sign test is significant for Exp. I and III at $p < .001$.

left sides of a white paper of the same reflectance as the target and subtending the same visual angle as the target was 165 ftc. when placed flat behind the partition so as to intercept the light of the projector. When seen in its true position, the target was seen to stand in the general illumination provided by the overhead lamp. Thus, in Exp. III and IV, both the illuminated and shadowed sides of the target were seen to be much more strongly illuminated when the target was apparently flat.

The *O*s were instructed to judge the lightness of the shadowed side of the target in Exp. I and III, and of the illuminated side of the target in Exp. II and IV. Each *O* was told that he would be given two looks at the target. His task was to note the lightness of the indicated side of the target and on the second look to match it to one of the grays on the Munsell scale. Each look was approximately 10 sec. All *O*s viewed the target first binocularly and then, after an interval of 20–30 min., monocularly. Twelve *O*s served in Exp. I and III, and 10 *O*s in Exp. II and IV. All *O*s were undergraduates who were paid to participate and were naive as to the purpose of the experiments.

Results and Discussion

Table 1 presents the medians of *O*s' Munsell matches in each of the four experiments. The results show that the shadowed side of the target is judged darker when the target is apparently flat than when the target is seen in its true position. The difference is .5 of a Munsell step in Exp. I and 1.25 Munsell steps in Exp. III (sign tests $p < .001$). A sign test of the increased lightness difference in Exp. III is significant at the .01 level. The *O*s' lightness judgments of the illuminated side of the target in Exp. II and IV, however, were not changed with changes in apparent spatial position. Regardless of apparent position, *O*s' lightness matches in Exp. III and IV with the added luminance of the projector were below those in Exp. I and II with only the overhead light (sign test $p < .001$). Thus, the results indicate that the perceived lightness of the target was affected by both the luminances in the field and the apparent position of the surface.

The question posed is: How do cues for spatial position influence the perception of lightness? The results fail to support the albedo hypothesis. The small changes in Exp. I and III indicate that perceived lightness does not vary directly with the apparent illumination. Had the target actually been located in the perceived position, the illuminance of the shadowed side of the target would have been increased by a factor of 1.5 in Exp. I and 58.9 in Exp. III. The failure of a lightness change to occur in Exp. IV indicates that a change in the apparent illumination is not a sufficient condition for a lightness change. Had the target actually been located in the perceived position, the illuminance of the illuminated side would have been increased by a factor of 39.3. Both the small lightness change of the shadowed side and the failure of a

lightness change to occur on the illuminated side are in agreement with the second hypothesis proposed. They suggest that the apparent position of the target is important in perceiving the area of reduced luminance either as a difference in surface lightness or as a difference in illumination. When the target is seen folded, the right side is seen as indirectly illuminated, and the area of lower luminance is perceived as an area of lowered illumination or as a shadow. However, when the target is seen as flat, perception of a shadow is no longer consistent with the apparent uniform illumination of the right and left sides of the target. The perception therefore changes to remove this discrepancy and the area of lower luminance is perceived as an area of lower surface reflectance, i.e., a darker surface color. Since no discrepancy, however, is introduced with the change of apparent position on the illuminated side, no lightness changes would be expected to occur.

The results, moreover, indicate that the range of lightness perceptions is determined by the luminance distribution reflected and not by the apparent illumination. If a surface is perceived to be illuminated uniformly, perceived lightness is determined by the relationship of luminances in adjacent regions, i.e., contrast. Thus, as would be expected on the basis of contrast, *O*s' lightness judgments with the added luminance of the projector in Exp. III and IV were below those in Exp. I and II. If a surface is perceived to be directionally illuminated, however, a change in perceived lightness may occur as a result of a new organization of the stimulus pattern.

Effects of set.—The previous discussion suggests that whether an area of lower luminance will be perceived as a shadow or as a darker surface color may depend on *O*s' readiness to infer from the available cues that the reduced luminance is the consequence of a reduced illumination or of a reduced surface reflectance. This possibility was able to be tested in terms of an unexpected observation. With sophisticated *O*s, the shadowed side of the target was judged darker independently of whether *O*s viewed the target monocularly or binocularly first. With unsophisticated *O*s, however, the order of observation proves to be crucial. If *O*s first see the target as flat (monocularly) and then folded (binocularly), no consistent changes of lightness occur. In one experiment of 12 *O*s, 5 *O*s judged the shaded side of the target to be darker when seen as flat, 5 *O*s judged it to be darker when seen folded, and 2 *O*s judged the lightness to be the same. In a second experiment of 11 *O*s, 3 *O*s judged the shaded side to be darker when seen folded, 4 judged it darker when seen flat, and 4 judged the lightness to be the same. The effect of order of observation indicates two important points. First, the results are again not consistent with the albedo hypothesis. Secondly, they rule out an explanation of the lightness changes as due to binocular interaction.[3] According to either of these hypotheses, the order of observation should make no difference. The observed asymmetry, however, is compatible with the second hypothesis proposed. If the target is initially seen as two gray shades, seeing it in its true position introduces no marked in-

[3] The hypothesis of binocular interaction is also ruled out by the fact that lightness changes will occur with changes in the apparent position of a target without changing from monocular to binocular observation, vid. Mach (1959, p. 209), Katona (1935), and Hochberg and Beck (1954).

consistency to induce a reorganization of the percept and no consistent changes of lightness would, therefore, be expected to occur. A lightening of the shaded side may be expected to occur with binocular observation following monocular observation if a preparatory set induces *O*s, when the target is seen in its true position, to perceive the area of lower luminance as an area of lower illumination. Such a set was established in two ways.

In one experiment, *O*s viewed the target first, binocularly-monocularly, and then after 10 min., viewed the target monocularly-binocularly. The expectation was that by first seeing the target binocularly-monocularly the presumption would be strengthened to see the right side of the target as shaded when the target was seen in its true position following monocular observation. Only the overhead light was on and an interval of 20–30 min. separated each of the binocular-monocular judgments. Thirteen *O*s were run. Eight of the 13 *O*s on seeing the target binocularly after seeing it monocularly judged the shaded side of the target as lighter, while 5 reported the lightness of the target to remain the same. A further experiment indicated that consistent lightness changes could be obtained by providing *O*s with greater experience than one trial. Five *O*s who had viewed the target binocularly-monocularly a number of times all judged the shaded side of the target as lighter when viewing the target binocularly following monocular observation. With one trial, however, only a short interval should elapse between the binocular-monocular judgments, since in an initial experiment little or no success was obtained with a day interval. In a second experiment set was introduced by means of instructions. The instructions were designed to induce an object-directed attitude when viewing the target binocularly. When the target was seen in its true position, the instructions emphasized that the target was a spotted card which had been folded in its center like a bookcover. Forty-five minutes separated the monocular and binocular judgments. Fifteen *O*s were run. Eleven *O*s reported the target as lighter when seen in its true position, and 4 reported the lightness of the target to remain the same.

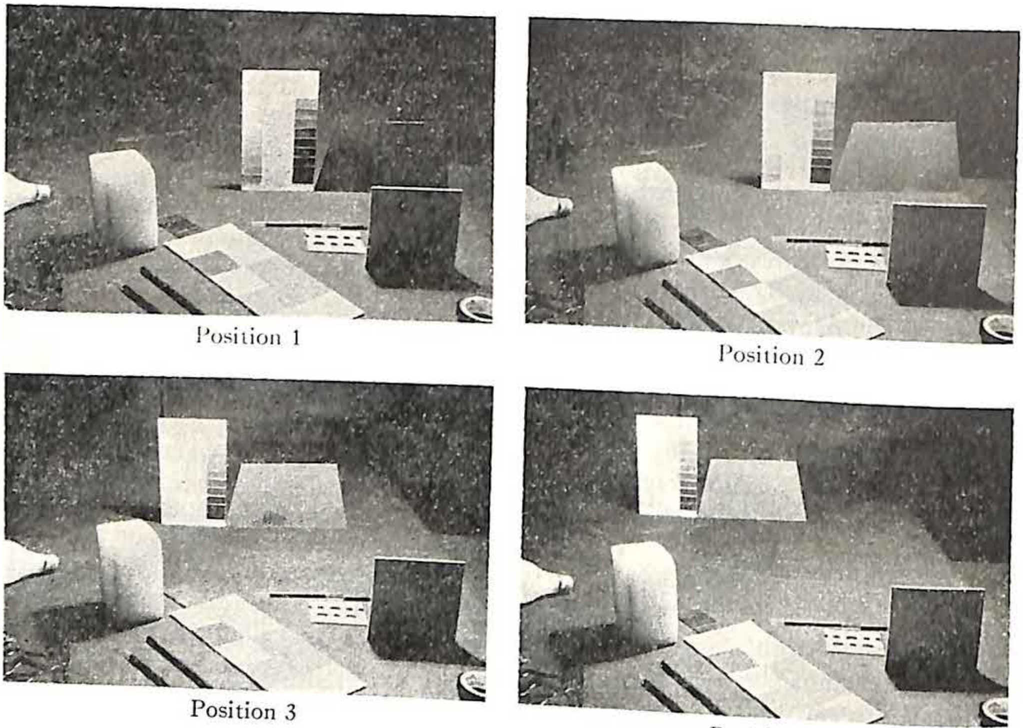
Study II

The effect of apparent spatial position on perceived lightness was also investigated using the experimental arrangement of Hochberg and Beck (1954). They have reported that a fixed target under constant illumination is judged darker when made to appear perpendicular to the apparent direction of illumination than when made to appear parallel to the direction of illumination. In a number of preliminary experiments, the position of the target was found to determine whether *O*s' lightness judgments were altered with changes in apparent position. The first investigation was designed to examine the effect of target position.

Method

The target, a trapezoid, was cut so that when placed upright its retinal image would be the same as that of a rectangle lying on a table surface. The trapezoid was covered with a dark gray paper having a reflectance of approximately 9.4%. The *O*s viewed the target through apertures cut in a screen fixed to one end of the table. These allowed *O*s to observe the target with either monocular or binocular vision. The target was seen as an upright trapezoid when cues for slant were present, and as a horizontal rectangle when cues for slant were absent.

The target was placed in four positions as shown in Fig. 2. The distances from the aperture to the center of the target were ap-



Position 1

Position 2

Position 3

Position 4

FIG. 2. Photographs of the table arrangement in Study II. (The target is shown in each of the four positions.)

proximately 46 in. in Position 1, 48 in. in Position 2, 50 in. in Position 3, and 57 in. in Position 4. The illumination was provided by an adjustable lamp attached to the table so that the light fell from above and behind Position 1. Ample cues were present to indicate the direction of illumination and its distribution. Changing the position of the target had two effects. First, it altered the distribution of light and shade over the surface. In Position 1 the target was heavily shadowed; in Positions 2 and 3 the degree of shadowing was decreased; in Position 4 the target did not appear shadowed at all. Secondly, the amount of change in apparent illumination resulting from changes in the apparent position of the target varied. The change of the target from apparently upright to apparently horizontal altered the perceived illumination most strongly in Position 1. The illuminance of the central area of the target was 13.5 ftc.; the illuminance of the central area of a gray paper of the same reflectance as the target placed flat on the table to correspond to the target projection was 140.6 ftc. The illuminances of the target and of a gray paper of the same reflectance as the target placed to correspond to the target projection on the table were 31.3 ftc. and 125 ftc., respectively, in Position 2; 56.3 ftc. and 93.8 ftc., respectively, in Position 3; and 62.5 ftc. in Position 4. Thus, if the target were actually a rectangle flat on the table the illuminance of the surface would be increased by a factor of 10.4 in Position 1, a factor of 4 in Position 2, a factor of 1.7 in Position 3, and would not change in Position 4.

Forty *O*s participated in the experiment; 10 *O*s judged the lightness of the target in each of the four positions. The *O*s were naive regarding the purpose of the experiment and were paid to participate. Each *O* made two judgments. He matched the lightness of the target to one of the grays on the Munsell scale when it was apparently upright and when it was seen to lie flat on the table.

Results and Discussion

The medians of *O*s' lightness matches in the four positions are shown in Table 2. Though in Positions 1, 2, and 3 the many cues for the

TABLE 2

MEDIANS OF *O*s' LIGHTNESS MATCHES OF THE TARGET PERCEIVED UPRIGHT AND HORIZONTAL

Positions	Rectangle	Trapezoid
1	3.5	4
2	4	4.5
3	5.25	5
4	5.5	5.5

Note.—In Position 1 lightness change was significant at $p < .01$.

illumination would indicate the target to be more strongly illuminated when apparently flat, a significant lightness change occurred only in Position 1. The median change in Position 1 was .5 of a Munsell step. The results confirmed again that a change in apparent illumination is not a sufficient condition for a change in apparent lightness and that apparent lightness does not vary as a linear function of the apparent illumination.

The results pose the question of why there was a consistent change of lightness with a change of apparent position in Position 1 but not in the other positions. The answer to this question would appear to be tied to a more basic question. The targets did not reflect light uniformly, but a distribution of luminances. This distribution, however, yielded a specific perception of surface lightness. How did this occur? A hypothesis consistent with what has been proposed is that a specific lightness perception results from an organization of the stimulus pattern in terms of the cues provided by the character and distribution of the luminance variations. That is, due to the shape, size, edge gradients, and distribution of the luminance variations, luminances differing from a particular value are seen as shadows, highlights, and light spots rather than as surface lightness differences (Beck, 1964). Thus, parts of the target were perceived as evenly illuminated, parts as shadowed, and parts as highlighted. As with the Mach card, the change of lightness in Position 1 demonstrates that this organization of the stimulus pattern is not unique but can depend on the apparent orientation of the surface. In general, however, whether a change in apparent position induces consistent changes in surface lightness, may be expected to depend on the pattern of luminance variations over the surface. For example, a preliminary experiment suggests that whether *O*s reported the target darker or lighter when apparently flat in Position 2 depends on whether they tended to fixate a shadowed or highlighted area of the target when it was seen upright. When told to fixate the upper portion of the target (which when seen upright was seen as highlighted), 8 of 10 *O*s judged the target when apparently flat to be lighter than when apparently upright. When told to fixate the lower portion of the target (which when seen upright was seen as shadowed), 7 of the same 10 *O*s judged the target when apparently flat to be darker than when apparently upright. A binomial sign test of the lightness changes due to the change in fixation is significant at the .02 level. The consistent darker judgments when the target was apparently flat in Position 1 may reflect the fact that the target in this position was seen to be almost entirely shadowed when apparently upright. Unfortunately, data are not available relating *O*s' pattern of fixations to the lightness changes which occurred with changes in apparent position.

Effects of contrast with background.—To separate the effects of contrast from the effect of cues for spatial position and illumination direction,

TABLE 3

MEDIANS OF *O*S' LIGHTNESS MATCHES OF THE TARGET PERCEIVED UPRIGHT AND HORIZONTAL WITH DIFFERENT BACKGROUNDS

	Rectangle	Trapezoid
Black background	4.25	4.75
White background	3.5	4

Note.—A Friedman two-way analysis of variance is significant at the .001 level.

an experiment was conducted in which *O*s judged the lightness of a target with a black and a white background. Because of contrast, regardless of its apparent orientation, the target on a white background should be judged darker than on a black background. On the other hand, the target should be judged on both a black and a white background as darker when perceived perpendicular to the direction of illumination than when perceived parallel to the direction of illumination. The experimental arrangement and procedure were similar to that of the previous experiment. The incident illumination was set to come from overhead. The illuminance of the central area of the target was 30.2 ftc. The illuminance of a gray paper of the same reflectance placed flat on the table to correspond to the target projection was 62.5 ftc. Ten *O*s were run who judged the lightness of the target with both a black and a white table cover. A balanced order was employed with half of the *O*s viewing the target on each background first.

Table 3 presents the medians of *O*s' lightness matches. The results indicate that the target was judged darker when apparently flat than when apparently upright, independent of whether it was on a black or a white background. Similarly, *O*s judged the target on a white background to be darker than on a black background independent of the orientation of the target. A Friedman analysis rejects the hypothesis that lightness matches at the different apparent positions and backgrounds differed only by chance. Sign tests of the differences of *O*s' lightness judgments due to the apparent position were significant at the .01 level for both a black and a white background; the differences due to the background were significant at the .01 level at both apparent orientations. Thus, the results separate the effects of contrast from the effects of apparent spatial position and illumination direction on perceived lightness.

DISCUSSION

The results of these experiments support earlier findings that the cue properties of stimuli as well as the luminance relationships affect lightness perception. The data, however, are inconsistent with the hypothesis that surface lightness is the result of computing surface albedo. Rather, the general hypothesis the studies appear to support is that certain processes of organization come into play as a result of the cue properties of stimuli which affect whether a variation in luminance will be seen as a difference in the illumination of the surface or as a difference in the lightness of the surface. Though going beyond the scope of the data, a brief discussion of two questions relating to the hypothesized organizational process is useful.

The hypothesized organizational process is assumed to be of central origin involving the integration of differentiated elements into a systematic whole. A characterization of the organizational process must, among other things, specify the characteristics of the stimulus pattern which is centrally integrated. Studies of lightness perception indicate that the central organization is not directly of the luminance pattern, but occurs only after the operation of more peripheral sensory processes, such as the processes of contrast and adaptation, processes of contour formation and surface formation

(Wallach, 1949). What is suggested is that the output of these more peripheral processes, i.e. a "lightness pattern,"[4] constitutes the input to the presumed organizational process. Thus, though the perception of surface lightness cannot be brought into a one to one correspondence with the luminance distribution, the luminance distributions through the operation of the peripheral sensory processes limit the range of surface lightness perceptions. Moreover, through the determination of edge characteristics and mode of appearance, i.e., whether a lightness gives the appearance of a film or a dense surface, the perception of shadow and light is influenced greatly by the luminance distribution. A second question concerns the way in which different factors operate to produce an organization of the stimulus pattern. The organization of the stimulus pattern is a complex relational affair involving the interaction of multiple factors, such as cues relating to the stimulus pattern; cues relating to the illumination and apparent spatial position of a surface; the factors of memory, set, and attitude (Woodworth & Schlosberg, 1954, Ch. 15); as well as certain organizational tendencies (Attneave, 1954). Exactly how these factors cooperate in achieving an integration of the stimulus pattern determined by the peripheral sensory processes is not known. The organization of the stimulus pattern, however, would appear to involve more than a response to the potency of individual cues. The experiments in Study I, for example, suggest that there exists a presumption to see a surface as being a single lightness if the conditions are consistent with the variations of luminance being seen as shadows and highlights. A convenient working hypothesis is probably to assume that the organization of a lightness pattern is such as to maximize the congruences between the different cues and to provide an economical summary of the stimulus pattern.

[4] The term "lightness pattern" is not intended to imply that there is an awareness, i.e. sensations, which corresponds to the neural patterns resulting from the peripheral sensory processes. The term is used solely to characterize through analogy the information contained in the neural pattern (that is, the neural excitations resulting from the luminance distribution) prior to organization.

REFERENCES

ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.

BECK, J. Judgments of surface illumination and lightness. *J. exp. Psychol.*, 1961, **61**, 368–377.

BECK, J. Supplementary report: An examination of an aspect of the Gelb effect. *J. exp. Psychol.*, 1962, **64**, 199–200.

BECK, J. The effect of gloss on perceived lightness. *Amer. J. Psychol.*, 1964, **77**, 54–63.

HELSON, H. Some factors and implications of color constancy. *J. Opt. Soc. Amer.*, 1943, **33**, 555–567.

HOCHBERG, J. E., & BECK, J. Apparent spatial arrangement and perceived brightness. *J. exp. Psychol.*, 1954, **47**, 263–266.

KATONA, G. Color contrast and color constancy. *J. exp. Psychol.*, 1935, **18**, 49–63.

MACH, E. *The analysis of sensations.* New York: Dover, 1959.

STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, **133**, 80–86.

WALLACH, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, **38**, 310–324.

WALLACH, H. Some considerations concerning the relation between perception and cognition. *J. Pers.*, 1949, **18**, 6–11.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received January 4, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 180–185

FACILITATION OF COMPETING RESPONSES AS A FUNCTION OF "SUBNORMAL" DRIVE CONDITIONS [1]

CHARLES Y. NAKAMURA AND WILLIAM E. BROEN, JR.
University of California, Los Angeles

The effect of very low drive states on the probabilities of a dominant and a competing response were investigated. The response hierarchy was established through training procedures with dominant and competing responses having different reinforcement probabilities. Ss were 40 male, college students. The results were consistent with the hypothesis, based on certain constructs from Hull-Spence theory, that competing responses are facilitated by reduction of drive within the range of low drives. It was concluded that the results support existing views on the important role of low drive in the modification of behavior.

A generally recognized assumption, particularly in the areas of counseling and psychotherapy, is that, when behavior patterns have been well learned, behavior change is facilitated by establishing a low, or subnormal, drive state (relaxed, low-anxious) in the individual. This is implemented by techniques such as the use of the couch, the creation of good rapport, making the person comfortable, and the like. Wolpe (1958) has been one of the most explicit users of this assumption. In the interest of modifying strong dominant anxiety responses, he has used specific relaxation procedures. One way of looking at procedures like the above where relaxation has been used is to view relaxation as a response which competes with the dominant response, anxiety. Another more general way of seeing the role of relaxation in behavior change is to view relaxation, not necessarily as a response, but as a lowering of a general drive state through which the relative superiority of any dominant response would be reduced and any competing response would be facilitated. If "subnormal" drive in this latter sense is effective in increasing the probabilities of nondominant responses, the use of relaxed states would have general importance for many areas where behavior change is desired. Examples are the areas of response originality (Maltzman, 1960) and creativity (Mednick, 1962), where theoretical discussions have emphasized the availability and use of nondominant responses.

The purpose of this experiment was to investigate the effect of lower than "normal" drive (very relaxed) states on the probabilities of a dominant (.8 reinforcement probability) and a competing (.5 reinforcement probability) response. The *S*s were initially trained in this two-response situation under conditions usually designated as low drive. It was hypothesized that further reduction in drive below an already low drive state would produce an increment in the response which has the lower probability of reinforcement, with a corresponding decrement in the response which has the higher probability of reinforcement. This hypothesis was contingent on the assumption that the

[1] The authors wish to thank Fred Shima for running *S*s and Steve Young for setting up the apparatus. This research was supported by funds from the University Research Committee.

response strength of the competing response would remain above threshold even in the low drive condition.

The hypothesis is consistent with certain constructs from Hull-Spence theory (Spence, 1956). According to the multiplicative relation between drive (D) and habit strength (H), the evocation probabilities of the competing response should increase with reduction in D. This follows from the assumption that the probability of a response is a function of the difference in the response strengths between the dominant response (E_d) and the competing response (E_c). The magnitude of the response strength superiority of the dominant response over the competing response is seen as a direct, multiplicative function of D, $E_d - E_c = \mathrm{f}[D(H_d - H_c)]$. Thus, a decrease in D should be accompanied by a decrement in the probability of the dominant response and a relative increase in the probability of the competing response. If these hypothesized changes in response probabilities are found in the present experiment they would be considered to be a function of the lower drive since reinforcement conditions remained the same both during training under low drive and testing under still lower drive.

Method

Subjects.—The *S*s were 40 introductory psychology class male students. Ten *S*s were assigned to each of the four treatment groups described later.

Apparatus.—The *S* and the control equipment were in adjacent rooms separated by a one-way vision window. The *S*'s room contained an interviewing couch with wedge shaped headrest, a finger response unit, a speaker, and a microphone.

The finger response unit was anchored on a small bench beside the couch. The unit comprised a block of wood with a $2\frac{1}{2} \times 5\frac{1}{2}$ in. surface covered by a stainless steel plate. In the central region of the surface were two $\frac{3}{8}$-in. diameter holes spaced $1\frac{1}{2}$ in. apart on center. Each hole contained a finger contact pin that projected up to the level of the steel plate. An adjustable hand-forearm rest that extended from the side of the block could be adjusted for the angle of *S*'s arm as it rested in relaxed position along the couch, and for the length of forefinger. In each trial interval, simultaneous finger contact with a pin and the steel plate activated a transistor relay circuit which operated impulse counters.

Starting with forefinger on one pin, *S* was required to make one of two responses in each trial, either move his finger to the other hole (Move response) or not move his finger at all (Stay response). Total trials and frequencies of each type of response were recorded on counters.

Procedure.—The *S*'s room was dimly lighted by a 25-w. bulb. Illumination was just sufficient for *E* to observe and tabulate gross motor movements which provided one check on *S*'s state of relaxation.

The *S* was asked to lie on the couch and instructed to learn to make the correct response following each signal to respond (an approximately $\frac{1}{10}$-sec. duration, high-frequency tone). This would be followed by another signal ($\frac{1}{5}$-sec. duration, low-frequency tone) if his response was correct and no tone if incorrect. (However, in fact, one response was randomly reinforced .8 of the time and the other was randomly reinforced .5 of the time.) The interval from the respond signal to the point where reinforcement might be given was 3 sec., and from that point to the next respond signal was 2 sec.

The four groups of *S*s are distinguished by the response (Move or Stay) which had the low probability of reinforcement and was thus established as the competing response, and by the treatment given between the training and the test sessions. For two of the groups Stay was the competing response and Move was the dominant response, and conversely for the other two groups. Because the focus of the experiment was on changes in probability of occurrence of responses which are hierarchically ordered, a learning criterion was used to select *S*s with the .8 reinforced response clearly dominant by the end of the training trials. Only those *S*s were used who attained a criterion of 70% or greater dominant responses (30% or less competing responses) over the last 25 training trials. Seventy of 110 initial *S*s did not meet the training criterion.

The training procedures and response criterion that resulted in the rejection of the large number of *S*s perhaps requires comment. The .5 and .8 reinforcement probabilities were

chosen with the objective of establishing a stable dominant response, yet not so superior in habit strength that the probability of an increment in the competing response would be negligible under the amount of drive reduction used in the experiment. Using these reinforcement probabilities, the criterion of 70% dominant responses over the last 25 of 100 training trials was too severe. Of the first 60 *S*s only 15 reached criterion. When the procedure was changed to give 150 training trials 25 of 50 *S*s reached criterion. The proportions of *S*s reaching criterion by each of the two procedures were nearly identical in the experimental and control groups.

Within each pair of groups that had the same competing response, one group received a drive reduction (relaxation) treatment between training and test sessions. The other group (control) was given a nonrelaxation experience. Thus the four groups are designated as: Relax-Move (i.e., relaxation treatment with Move being the competing response), Relax-Stay, Nonrelax-Move, and Nonrelax-Stay.

The drive-reduction procedure consisted of telling *S* over the speaker to follow instructions for four relaxation exercises recommended by Jacobson. Slightly modified from the original (Jacobson, 1938), these instructions were spoken slowly and with pauses to pace *S*'s movements:

1. Raise both arms straight up. Clench your fists. Hold very tightly. Now slowly let your arms fall to your sides and go completely limp.
2. Now push with your heels straight out toward the wall, hard, and as far as you can. Hold this position tightly. Slowly let your legs relax against the cot.
3. Now take a deep breath—and when you exhale let your chest drop completely.
4. Next wrinkle up your forehead as tight as you can, raise your eye brows as high as possible. Then gradually let go and relax your face completely. Relax your forehead and your eyes. Relax your mouth and jaws—completely relax.

The control groups were asked to solve a problem mentally; to take the number 10,000, divide by 4 until the smallest possible quotient was obtained, and when an answer was derived to recheck it. Following the respective instructions all *S*s were told that the task in the second session was to be identical to that in the first. After the test trials, each *S* rated his degree of relaxation in each of the preceding two sessions on a 7-point scale ranging from Not relaxed to Very relaxed. A self-rating anxiety inventory (Nakamura, 1960) was administered at the end of the experiment.

The main stages in the procedure may be summarized: 100 or 150 training trials, 4-min. interval during which the two experimental groups received relaxation exercises and the two control groups mentally solved the arithmetic problem, 100 test trials, self-ratings on relaxation scale, and anxiety inventory. The reinforcement conditions remained the same over both training and test trials.

Results

The mean percent competing responses in the last 25 trials of training are shown in the left side of Fig. 1. The analysis of variance applied to the mean scores indicated that differences between the Relaxation Treatment and Response Type groups, and the Treatment × Response Type interaction were all nonsignificant. Thus the four groups had attained essentially the same response strength of the competing response at the final stage of training. A further check of the comparability of groups on training response strengths was made by comparing the Relax and the Nonrelax groups on percent competing responses in the 25 trials preceding the 25 criterion trials. This difference was also clearly nonsignificant.

The mean percentages of the competing response in blocks of 25 test trials are presented by the connected points in Fig. 1. Percent competing responses for the two Nonrelax groups remained about the same level over blocks of trials. In contrast, the curves for the two Relax groups show increments over blocks.

An analysis of variance using the difference scores between the mean of the last 25 training trials and the mean of the 100 test trials yielded a significant Treatment effect, $F(1, 36) = 8.21$, $p < .01$. This confirmed the apparent differences shown by the

curves in the amount of increase in percentage of competing responses between the Relaxed and Nonrelaxed Treatment groups. The Response Type effect and the Treatment × Response Type interaction effect were not significant.

A similar analysis of variance applied to the difference scores between the last 25 training trials and the first 25 test trials yielded a significant Treatment effect, $F(1, 36) = 4.40$, $p < .05$. An analysis of the difference between the last 25 training trials and Test Trials 76–100 was also significant, $F(1, 36) = 12.49$, $p < .01$. Together, these results indicated that there was a slight but significant immediate effect of the Treatment and that this effect was maintained to the end of testing. The effect of the Treatment appeared to increase over trials. However, a trend test of the linear component yielded nonsignificant results. The Treatment effect did approach significance, $F(1,36) = 3.81$, $p < .10$.

Of the secondary measures obtained in the experiment, only the subjective ratings by *S*s on the relaxation scales approached significance, $F(1, 36) = 3.72$, $p < .10$. The *S*s in the Relaxation Treatment groups tended to rate themselves more relaxed in the test-trials session than in the training session while the nonrelaxation groups showed no measurable changes. Parenthetically, it may be mentioned that the measurable amount of change was limited by a ceiling. The relaxation ratings of the first session were already very high with an overall mean of 4.8 on a 7-point scale. Analyses of scores for the remaining two measures, *E*'s rating of *S*'s movements, and the anxiety inventory, were nonsignificant.

All *S*s used in the experiment, including those that were rejected, were

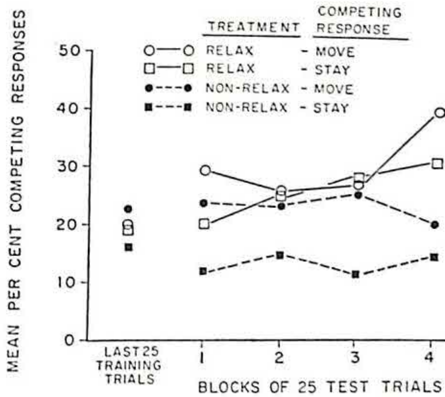


FIG. 1. Mean percent competing respons in the last 25 training trials, and over fo blocks of test trials for four groups of 10 *S*s

randomly assigned to the Relax an Nonrelax treatments and given te trials as well as training trials sin there was not time to score f criterion performance level immed ately following the training sessio Thus, test data were available for th 70 *S*s who were rejected because the did not meet the training criterion 70% dominant responses. Inspectio of these data showed that the mea response levels of the .8 reinforcemen probability response in the last 2 training trials were 48% and 49% respectively, for *S*s assigned to th Relax and Nonrelax treatments (3 *S*s in each group). The means of th .8 reinforcement response over th total test trials were an identical 53% for both groups; the response leve plotted over four 25-trial blocks bein generally flat with a very slight in crease over blocks. Thus, the overa performances of the rejected *S*s as signed to the two treatments wer quite similar. The data sugges furthermore, that *S*s in whom dominant response is not establishe in 100 or 150 training trials with th procedures used in this experiment ar not likely to attain a clearly dominar

response in an additional 100 trials of a second session.

Discussion

The results were consistent with the hypothesis that competing responses are facilitated by reduction of drive at relatively low drive levels. In the response with the lower probability of reinforcement, which was established as the competing response, the relaxation procedure produced an increment in that response as compared to the nonrelaxed condition. The results are also consistent with the Hull-Spence hypothesis of decrease in drive leading to decrease in $E_d - E_c$ in conditions where response strength of the competing responses is above threshold. Where drive levels are lower than those used in this experiment, or where the strength of the competing response is less, it is possible that a reduction in drive would cause competing response strengths to go below threshold and the facilitation of competing responses found in this study may not be observed.

There are, of course, alternative explanations that might be put forth to account for the obtained effects. These are inadequate attention to the task by the relaxed *S*s, and different drive stimuli (S_D) associated with the different drive levels at testing.

Very possibly the relaxation procedures could reduce *S*'s attentiveness to the task. However, if attention were inadequate owing to sources such as reduced concern about performance or *S*'s bordering on falling asleep, this would affect most the condition where the dominant response was the Stay response than when it was the Move response. There was no significant difference found between these conditions as indicated by the nonsignificant Response Type effect. Moreover, the presence of at least a minimally adequate attention level was indicated by the fact that trials in which a response could not be distinctly scored as a Move or a Stay response were a rarity and those that did occur were almost equally scattered over the four groups.

The criticism that the experiment did not control for possible differences in S_D at the different drive levels does not appear to be particularly critical for interpretation of the results since the conditions of reinforcement remained the same for each *S* throughout both the training and test sessions. With this procedure, responses early in the test trials would be affected most by changes in S_D, and the effect would be expected to diminish as relearning under the new S_D occurred with continued trials. This is because the reinforcement probabilities would become effective with the continued trials and would be expected to produce a recovery and possibly an increase in the strength of the dominant response. The present results, on the contrary, show a tendency for a continuing decrement in strength of the dominant response. This explanation would pertain whether S_D changes were attributed to changes in drive level through relaxation exercises or through the possibility that *S*s interpreted the task in the test session to be different in some way in spite of having been told by *E* that it was to be "identical."

The further question of whether verbal instruction alone would have been as effective as the complete relaxation procedure used is indicated by evidence that verbal instruction or information can reduce the strength of a conditioned response as a function of reduced drive (Lindley & Moyer, 1961; Wickens, Allen, & Hill, 1963). However, these studies involved considerably higher drive levels than in the present experiment. It seems reasonable to expect that in this experiment the effects of verbal instruction and actual physical relaxation combined to produce low drive. A tentative indication that *S*s given the relaxation treatment felt more relaxed in the second than in the first session, and that there was no measurable change in the control groups was given by the data from the subjective ratings of relaxation. The available data are not adequate to satisfactorily answer the question of the

relative effectiveness of verbal and physical factors. This still remains to be investigated.

Several studies have reported that reduction in drive produces a decrement in a response (Grings & Lockhart, 1963; Lindley & Moyer, 1961; Wickens et al., 1963). Others have demonstrated that relaxed conditions yield a greater variability of responses to stimuli (Horton, Marlowe, & Crowne, 1963; Siipola, Walker, & Kolb, 1955; Usdansky & Chapman, 1960). However, most of this work was not concerned with establishing an especially low level of drive. Furthermore, little or no emphasis has been given to the specification and control of the competing response, which is of central importance if the primary concern is the facilitation of such responses. The present study differed from previous research in its attempt to build in specific dominant and competing responses of given strengths and to investigate the probabilities of these responses as a function of reduction of drive within the range of low drives. This experimental control gives a sound base for the conclusion that lower than normal drive states facilitate competing responses. The results thus provide support for the existing views on the important role of low drive in the modification of behavior.

REFERENCES

GRINGS, W. W., & LOCKHART, R. A. Effects of "anxiety-lessening" instructions and differential set development on the extinction of GSR. *J. exp. Psychol.*, 1963, **66**, 292–299.

HORTON, D. L., MARLOWE, D., & CROWNE, D. P. The effect of instructional set and need for social approval on commonality of word association responses. *J. abnorm. soc. Psychol.*, 1963, **66**, 67–72.

JACOBSON, E. *You can sleep well.* New York: McGraw-Hill, 1938.

LINDLEY, R. H., & MOYER, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *J. exp. Psychol.*, 1961, **61**, 82–88.

MALTZMAN, I. On the training of originality. *Psychol. Rev.*, 1960, **67**, 229–242.

MEDNICK, S. A. The associative basis of the creative process. *Psychol. Rev.*, 1962, **69**, 220–232.

NAKAMURA, C. Y. Measures of over-controlled and under-controlled behavior: A validation. *J. clin. Psychol.*, 1960, **16**, 149–153.

SIIPOLA, E., WALKER, W. N., & KOLB, D. Task attitudes in word association, projective and non-projective. *J. Pers.*, 1955, **23**, 441–459.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

USDANSKY, G., & CHAPMAN, L. J. Schizophrenic-like responses in normal subjects under time pressure. *J. abnorm. soc. Psychol.*, 1960, **60**, 143–146.

WICKENS, D. D., ALLEN, C. K., & HILL, F. A. Effect of instructions and UCS strength on extinction of the conditioned GSR. *J. exp. Psychol.*, 1963, **66**, 235–240.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford Univer. Press, 1958.

(Received January 6, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 186–192

UCS PROPERTIES IN CLASSICAL CONDITIONING OF THE ALBINO RABBIT'S NICTITATING MEMBRANE RESPONSE [1]

ALFRED BRUNER [2]

University of California School of Medicine, Los Angeles

The effects of light onset and corneal puff as UCSs were investigated. In Experiment I a group receiving 100% light UCS trials failed to condition, contrary to contiguity theory expectations, since UCRs were reliably elicited, while a 100% puff group showed clear conditioning. Another group, receiving 50% puff and 50% light UCSs (50/50-PL), was superior to a conventional 50% puff group (50-P), the light apparently mediating puff effectiveness. In Experiment II a group receiving 50% puff and 50% vibrotactile UCSs performed midway between Groups 50/50-PL and 50-P asymptotically. Light onset was not reinforcing, but instead depressed responding below the normal extinction level in a group switched from 100% puff to 100% light UCS trials. All 50% groups exhibited within-session decrements and between-session recovery.

Among classical conditioning (CC) studies relevant to the contiguity-reinforcement issue are those which have attempted to condition the pupillary reflex using as the UCS either a change in illumination or shock (Kimble, 1961, p. 52). These studies have shown that reliable pupillary CRs (dilation or constriction) are not observed when the UCS is a change in illumination, while reliable CRs (dilation) are observed when the UCS is shock. One such study suggested that the "reinforcement characteristic" of shock facilitates conditioning, while light, "unless the intensity is very high, is not considered to have intrinsic reinforcement properties [Gerall, Sampson, & Boslov, 1957, p. 468]." Since both shock and light offset elicit dilation UCRs, the use of either as the UCS in the presence of a neutral stimulus satisfies the essential conditions of the traditional notions of stimulus-response contiguity. Consequently the studies of pupillary conditioning are taken to support a reinforcement interpretation of CC.

The present studies parallel those on pupillary conditioning in that light was employed as a UCS and its effects compared to a known reinforcing stimulus (corneal puff). In the present instance the response under study was nictitating membrane extension in the albino rabbit, whose conditioning with a corneal puff UCS has been reported previously (e.g., Gormezano, Schneiderman, Deaux, & Fuentes, 1962).

Experiment I

Method

Subjects.—The Ss were 32 naive, male and female albino rabbits, 85 to 100 days old, procured from Windsor Biology Gardens,

[1] Based on a thesis submitted in partial fulfillment of the requirements for the PhD degree, Indiana University, 1963. The research was performed during my tenure as a United States Public Health Service predoctoral fellow and supported in part by National Science Foundation Grant GB-145 to I. Gormezano, to whom I wish to express my gratitude for his advice and assistance throughout the course of this research.

[2] Now at the University of Southern California.

Bloomington, Indiana. The animals were housed in individual living cages with free access to food and water.

Apparatus.—The apparatus, potentiometer-polygraph recording system, and method of restraint have been described in detail previously (e.g., Gormezano, 1965). Four *Ss* were run simultaneously, each being restrained in a Plexiglas holder within a ventilated refrigerator shell. Continuous indirect illumination was provided by a 6-w., 120-v. bulb located 1 ft. above *S*, adjacent to which was a 200-w., 120-v. incandescent lamp which, when illuminated, served as a UCS for some groups.

The CS was a 3,500-cps tone (72 db. SPL, .0002 ref.) presented through a speaker 5 in. in front of *S*. The puff UCS was compressed nitrogen under sufficient pressure to support an 80-mm. column of mercury as measured at the eye. Electronic timers were set to provide a CS duration of 900 msec. and a UCS duration of 400 msec. Onset of the UCS occurred 500 msec. after CS onset and the two stimuli terminated together. These timing specifications are accurate for trials on which the puff UCS was used. On light UCS trials, however, the CS-UCS interval was about 600 msec., rather than 500 as on puff trials, due to the relatively slow rise time of the 200-w. lamp, which was judged to become an effective UCS about 100 msec. after application of current. The intertrial intervals were 15, 25, and 35 sec. in a random order, with a mean of 25 sec. A CR was defined as an upward deflection of the response pen 1 mm. or greater from the record base line between 50 and 500 msec. after CS onset.

Procedure.—The *Ss* were divided into four groups of eight animals each. Group 100-P was a conventional CC group receiving a corneal puff as the UCS on all acquisition trials; Group 100-L received light onset as the UCS on all acquisition trials; Group 50-P, a conventional partial-reinforcement (PR) group, received a puff on 50% of the acquisition trials and no UCS on the other 50%; Group 50/50-PL received a puff UCS on 50% of the acquisition trials and the light UCS on the other 50%.[3] The randomized sequence of trials on which puffs were given was the same for Groups 50-P and 50/50-PL.

Following two adaptation sessions, during which spontaneous responses were recorded, all *Ss* received 70 acquisition trials per day for 10 consecutive days. On the day following completion of acquisition 200 extinction trials were given. On the next day 200 additional nonreinforced trials were given and this day will be designated spontaneous recovery.

Results

Since both the light and puff UCRs involved nictitating membrane extension (closure), their recorded topographies appeared quite similar, with the light UCR usually having a somewhat longer latency (100 msec. vs. 50 msec.) and slower rise time than the puff UCR. Neither UCR evidenced an opening (orienting) component, nor did the UCRs differ noticeably between groups or over trials.

All statistical analyses were performed on the arcsin transform of percentage CRs over 10-trial blocks. By convention, however, the figures are presented in terms of nontransformed percentage CRs. Figure 1 presents each group's mean percentage CRs plotted over 70-trial blocks during acquisition and 40-trial blocks during extinction and spontaneous recovery. During adaptation, responding did not exceed the spontaneous blink rate typical for rabbits (two or three per hour) and was therefore too low to be plotted clearly in Fig. 1.

Figure 1 indicates that Group 100-L does not appear to demonstrate any evidence of conditioning, although UCRs were reliably elicited by the light throughout acquisition. Group 100-L was consequently excluded from the data analyses to avoid error underestimation. An analysis of variance over all acquisition revealed significant differences between groups, $F(2, 21) = 5.21$ (the .05 significance level is utilized throughout the report), and individual *t*-test comparisons indicated significant differences between Groups

[3] In a pilot study a group of eight animals treated almost identically to Group 100-L had failed to develop CRs. Group 50/50-PL was therefore devised to provide a potentially more sensitive measure of the light's effects.

100-P and 50/50-PL (t = 2.17) and Groups 100-P and 50-P (t = 3.21; two-tailed t critical = 2.03 for both comparisons). The chance likelihood of the difference between Groups 50/50-PL and 50-P over the last two acquisition sessions was p = .06 (Fisher median test, Siegel, 1956).

On the right side of Fig. 1 are the performance curves for extinction and spontaneous recovery, where, as is often the case in extinction data, there was considerable individual variability. Thus although Group 50/50-PL's response strength appears superior during both these sessions, t tests revealed that the only significant difference occurred during spontaneous recovery and was between Groups 50/50-PL and 50-P, t (14) = 1.81. No adjustments were made for differences between acquisition asymptotes in these comparisons. As in some previous studies (e.g., Grant & Schipper, 1952), the PR group (50-P) shows a slight increase in responding from the end of acquisition to the first data point of extinction, while the 100% group (100-P) shows a rapid, marked decrease. Thereafter these two groups perform similarly. Although the extinction data are collapsed into 40-trial blocks the early extinction findings are not obscured, since the initial points shown also accurately reflect the approximate percentages observed when plotted over 5-trial blocks.

Figure 2 displays the within-session (10-trial blocks) acquisition performances of Groups 100-P, 50-P, and 50/50-PL. As is apparent the curves of 100-P tend to be U shaped over the last four acquisition sessions, while those of the PR groups, especially 50-P, tend to show a progressive response decrement as the session proceeds, following an initial high level of responding relative to the end of the preceding day. The analysis of variance over all acquisition revealed a significant effect for 10-trial blocks, F (6, 126) = 2.59, indicating that the within-session performance changes shown were reliable.

Analyses of variance on the final two acquisition sessions, extinction, and spontaneous recovery revealed no significant differences between groups in mean response latency. The reader

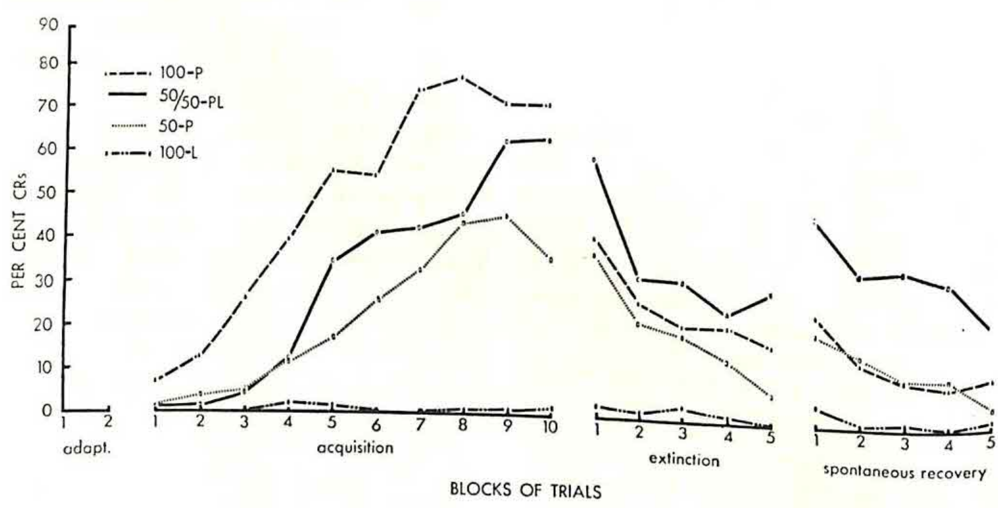


FIG. 1. Group mean percentage CRs for 70-trial blocks in acquisition and 40-trial blocks in extinction and spontaneous recovery.

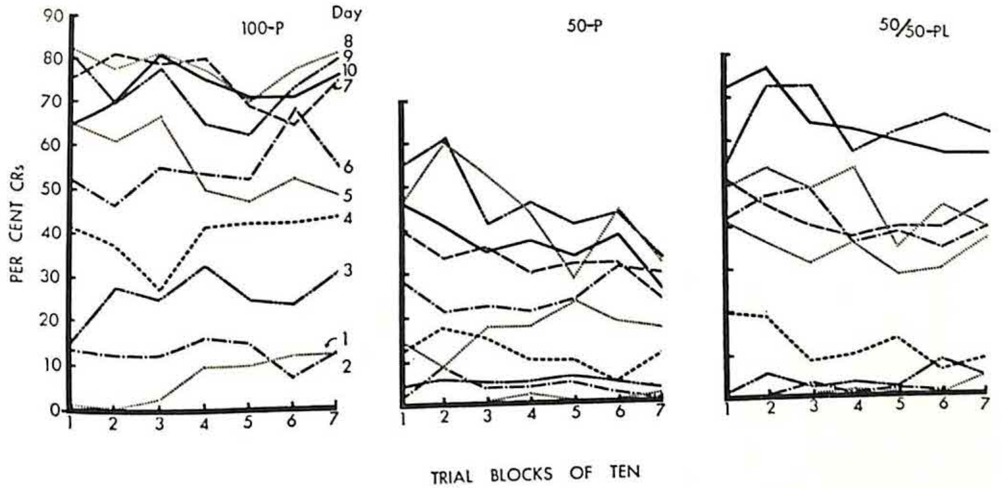


FIG. 2. Group mean percentage CRs over 10-trial blocks showing a separate curve for each of the 10 days of acquisition.

is referred to Bruner (1963) for further details of the CR and UCR latency analyses.

Discussion

Group 100-L's failure to condition corresponds to the finding of Gerall et al. (1957) mentioned earlier, in which contiguity between CS, light UCS, and UCR was not sufficient to bring about detectable CRs. In contrast when the present UCS was a corneal puff (or when Gerall's UCS was shock) CRs readily appeared. Both the present UCSs (light and puff) were clearly defined and reliable in occurrence. Had a more intense light been used CRs might have developed in Group 100-L, for when UCS intensity is increased so is strength of conditioning (e.g., Ross & Hunter, 1959). But it is evident that increasing the light's intensity may increase its aversiveness as well as its perceptibility. Accordingly, Watson's (1916) early claim of "some" pupillary conditioning seems plausible in view of his describing the UCS as a "very strong light." Unfortunately there appear to be no reported attempts to replicate Watson's study, which by present day standards was too brief and casual to constitute acceptable evidence.

The within-session decrements and between-session recoveries exhibited by the PR groups suggest an inhibitory effect of nonreinforcement which cumulated within sessions and dissipated between them. The cumulative decremental effects of PR have been described previously (e.g., Ross, 1959), but since seldom is more than one session employed there are very few previous indications of the dissipation effect (Grant, Riopelle, & Hake, 1950; Grant & Schipper, 1952; Humphreys, 1939). The latter three studies did employ a second session and did observe a rise in PR performance at the outset of the second session, suggesting a dissipation effect which was not demonstrated by 100% groups. It would therefore seem desirable to employ multiple sessions in future studies of PR inhibition.

While the light did not bring about conditioning when presented as the sole UCS in Group 100-L, it appeared to enhance responding when interspersed among puff trials in Group 50/50-PL. The latter outcome could have been the result of the interspersing procedure, per se, unique properties of the light, or both.

Experiment II

This experiment sought to extend the findings of Exp. I by testing (*a*) whether a stimulus which does not elicit membrane extension would facil-

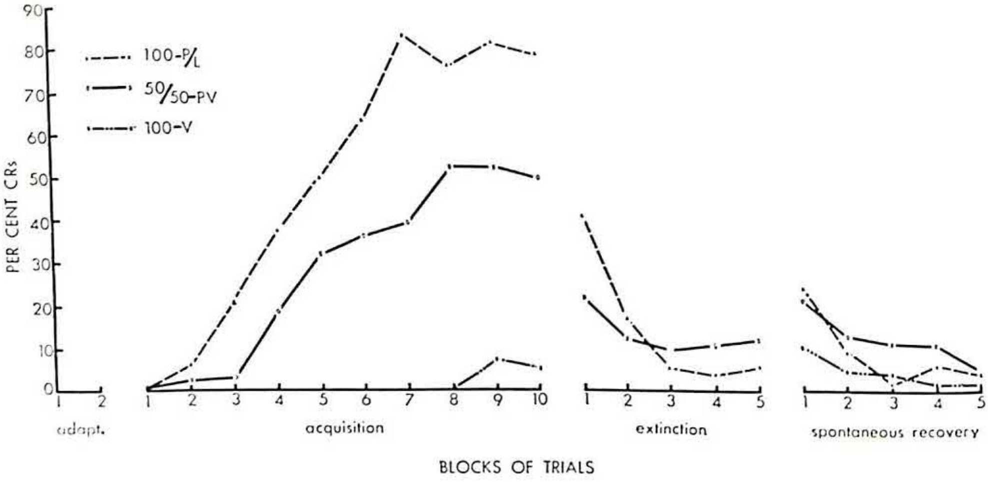


FIG. 3. Group mean percentage CRs for 70-trial blocks in acquisition and 40-trial blocks in extinction and spontaneous recovery for Exp. II.

itate conditioning when interspersed among puff trials as the light tended to, and (*b*) whether the light would maintain or enhance responding when it followed a series of puff trials as it had when interspersed among them.

Method

The method for Exp. II is identical to that for Exp. I except for the points noted below.

Subjects.—The *S*s were 20 albino rabbits.

Apparatus.—The only additional piece of apparatus used in Exp. II was a vibrotactile stimulator, 23 mm. diameter, 33 mm. high, 47.5 gm., constructed according to instructions by Bice (1961). Originally a small earphone assembly, the vibrator was firmly snap-fastened to the velcro strap passing over the animal's head allowing its 60-cps vibrations to be mechanically transmitted to the top center of the rabbit's head. Pilot work in which the vibrator was used as a CS indicated that it served as a clearly perceptible stimulus for the rabbit.

Procedure.—The animals were divided into two groups of eight each and one group of four. Group 100-P/L (eight *S*s) received 10 daily 70-trial sessions of 100% puff reinforcement during acquisition. The 200 extinction and 200 spontaneous recovery trials were modified for Group 100-P/L in that instead of presenting the CS alone, the light was introduced as a new UCS and the puff witheld. Group 50/50-PV (eight *S*s) received the puff on half the acquisition trials and the vibrator on the other half in the same sequence used for 50/50-PL earlier, and received conventional extinction (i.e. CS alone). Group 100-V (four *S*s) received the vibrotactile stimulus in place of a UCS on all acquisition trials followed by conventional extinction.

Results and Discussion

Figure 3 presents mean percentage CRs for each group plotted over 70-trial blocks in acquisition and 40-trial blocks in extinction and spontaneous recovery. During acquisition Group 100-P/L performed much like the previous 100% group both within (not shown) and over sessions (Fig. 3). Group 50/50-PV attained an asymptote at about the 50% level of responding, midway between the asymptotes of Group 50/50-PL (64%) and 50-P (40%) of Exp. I (cf. Fig. 1 and 3), and also manifested the within-session PR decrement. Group 100-V showed no indication of conditioning although one atypical rabbit was responsible for a few elevated points shown in Fig. 3. An analysis of variance over all acquisition for Groups 100-P/L and 50/50-PV revealed a significant F (1, 14) = 7.27 for groups.

On comparing the three PR groups

of both experiments, Group 50/50-PV's intermediate asymptote suggests that the interspersing procedure in itself facilitated conditioning somewhat, while the light as the interspersed UCS enhanced it even further. These effects are not clear statistically, however, as individual t-test comparisons failed to indicate ($.30 > p > .25$) that Group 50/50-PV was significantly different from either of the earlier PR groups over the last 2 days of acquisition. During extinction and spontaneous recovery, however, Group 50/50-PV's resistance to extinction was about the same as 50-P's (cf. Fig. 1 and 3) and significantly less than 50/50-PL's, t (14) = 2.09 for extinction, and 1.97 for spontaneous recovery, implying that unique properties of the light were responsible for the boosted response strength of Group 50/50-PL.

As contrasted to Group 100-P (Fig. 1), Fig. 3 shows Group 100-P/L's responding to be greatly depressed following the first 40 trials of "extinction" where the light replaced the puff as the UCS. The difference in performance of Groups 100-P and 100-P/L during the last 160 trials of extinction was compared by the Fisher median test (Siegel, 1956), which revealed a $p = .06$. On the final day in Fig. 3 Group 100-P/L again shows a reduced CR frequency following some spontaneous recovery at the initial data point.

Paradoxically there appear to be two conflicting findings with regard to the effects of the light UCS. When interspersed among puff trials the light tended to enhance response strength. But when presented as the UCS following puff conditioning the light markedly reduced CR frequency relative to regular extinction. To more closely examine the individual effects of the light, puff, and vibrator, a sequential analysis was performed on the data of the PR groups' last two acquisition sessions. This entailed frequency counts of CRs on each trial depending on the occurrence or nonoccurrence of a puff on the preceding trial. The difference in frequency of shifting from one response state (CR or no CR) to another, over pairs of trials, was corrected for overall response probability and utilized as an index (θ) of the average change in response probability over trials following administration of the particular UCS being examined.[4]

TABLE 1

THETA (θ) VALUES REFLECTING EXTENT AND DIRECTION OF CHANGES IN RESPONSE PROBABILITY FOLLOWING PUFF AND NONPUFF TRIALS

UCS Condition on Trial n	θ on Trial $n+1$		
	50/50-PL	50/50-PV	50-P
Puff	.236	.220	.085
No puff	−.123	−.208	−.124

Note.—The minus sign indicates a negative response tendency.

The θs reflecting changes in response probability on Trial $n+1$ as a function of whether or not a puff was given on Trial n are presented in Table 1 for the PR groups for the last two acquisition sessions. As the data of Table 1 suggest, the sequential analysis provided two main findings: (a) Following a puff on Trial n, response probability was boosted in all three PR groups on Trial $n+1$, but was clearly higher in the two 50/50 groups whose values were similar. (b) Following nonpuff trials, however, the three PR groups manifested a tend-

[4] The resultant ratio, θ, is commonly employed in mathematical learning experiments as a learning rate parameter. The complete sequential analysis and corresponding formulae are presented in Bruner (1964). Thanks are due to James J. Greeno for his assistance on this analysis.

ency to not respond, which was greatest in Group 50/50-PV with Groups 50/50-PL and 50/P being about equal. Together these findings suggest that the immediate effect of the substitute UCS was to depress responding (Finding *b*), while at the same time serving indirectly to support responding by somehow maintaining or mediating the puff's effectiveness at a level above that attributable to UCS omission (Finding *a*). Finding *a* indicates that puff effectiveness was maintained to an equal extent by both light and vibrator, implying that the properties of the substitute UCS did not constitute an important variable in producing the mediation effect. But the properties of the substitute UCS seem more important in the extent to which it detracts from response strength: the different amounts of response decrement following light and vibrator trials appears to be the major difference observed between these two stimuli (Finding *b*). The similar negative θs shown in Table 1 following light and omission trials (50/50-PL and 50-P, respectively) indicate that these two conditions produced equal amounts of decrement in the interspersing procedure. However, the light's decremental effect on the depressed "extinction" responding of Group 100-P/L was greater than that attributable to puff omission.

One hypothesis suggested by this discrepancy is that the interspersing procedure facilitated transfer (generalization) of stimulus properties from the puff to the light which was sufficient to mitigate the light's decremental effect somewhat, but was not sufficient to boost the light's θ up to a positive value. This transfer would not be expected to occur so readily in the interspersed vibrator group however, because the vibrator had fewer properties in common with the puff than the light had (e.g., both puff and light elicited membrane extension). Hence the decrement produced by the vibrator was not mitigated to the extent of that produced by the light.

REFERENCES

BICE, R. C. Electromechanical transducer for vibrotactile stimulation. *Rev. scient. Instrum.*, 1961, **32**, 856–857.

BRUNER, A. *Investigations of the properties of the unconditioned stimulus in classical conditioning of the nictitating membrane response in the albino rabbit.* (Doctoral dissertation, Indiana University) Ann Arbor, Mich.: University Microfilms, 1964, No. 64–5440.

GERALL, A. A., SAMPSON, P. B., & BOSLOV, G. L. Classical conditioning of the human pupillary dilation response. *J. exp. Psychol.*, 1957, **54**, 467–474.

GORMEZANO, I. Classical conditioning. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology.* New York: McGraw-Hill, 1965.

GORMEZANO, I., SCHNEIDERMAN, N., DEAUX, E., & FUENTES, I. Nictitating membrane: Classical conditioning and extinction in the albino rabbit. *Science*, 1962, **138**, 33–34.

GRANT, D. A., RIOPELLE, A. J., & HAKE, H. W. Resistance to extinction and the pattern of reinforcement. I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40**, 53–60.

GRANT, D. A., & SCHIPPER, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

HUMPHREYS, L. G. The effect of random alternation of reinforcement of the acquisition and extinction of conditioned responses. *J. exp. Psychol.*, 1939, **25**, 141–158.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

ROSS, L. E., & HUNTER, J. J. Habit strength parameters in eyelid conditioning as a function of UCS intensity. *Psychol. Rec.*, 1959, **9**, 103–107.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

WATSON, J. B. The place of the conditioned-reflex in psychology. *Psychol. Rev.*, 1916, **23**, 69–116.

(Received January 11, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 193–199

MONOPTIC AND DICHOPTIC VISUAL MASKING BY PATTERNS AND FLASHES [1]

PETER H. SCHILLER

Massachusetts Institute of Technology

This study investigated masking of letters by a bright flash of light or by a pattern. The results showed that (a) masking by flash is primarily a monoptic effect; masking by pattern occurs under monoptic and dichoptic conditions; (b) increasing the interstimulus interval decreases masking by pattern less than by monoptically presented flash; (c) repetition of trials decreases masking by pattern but not by flash. Different processes are involved when flashes and patterns are used as masking stimuli.

When two visual stimuli are presented in rapid succession, perceptual interference may result. This effect has been studied most extensively with respect to "backward" masking, when interference with perception of a first stimulus is studied as a function of the second stimulus in the pair (Raab, 1963). A number of mechanisms for this phenomenon have been proposed (Alpern, 1953; Boynton, 1961; Exner, 1868; Kolers, 1962; Stigler, 1910; Werner, 1935).

Several authors (Kolers & Rosner, 1960; Schiller & Wiener, 1963; Werner, 1940) have reported that perception of a visual pattern can be masked either monoptically (both stimuli to the same eye) or dichoptically (the two stimuli to separate eyes), when the masking stimulus is also a visual pattern. Masking also occurs when a visual form is followed by a stimulus having essentially no figural characteristics, such as a bright flash of light (Lindsley, 1961).

The aim of this study was to determine whether masking by pattern and masking by flash reflect identical processes or whether different mechanisms are involved. That the two might differ is suggested by the observation that masking by flash seems to be primarily a monoptic effect (Lindsley, 1961; Sperling, 1964), while masking by pattern has been shown to take place dichoptically. Accordingly, three experiments were performed. In the first, the extent of monoptic and dichoptic masking was investigated as a function of interstimulus interval in order to see if this temporal variable affects masking by pattern and flash in the same or different manner. In the second experiment, an analogous comparison was made with respect to the role of practice. In the third experiment, the flash and the pattern were presented together in various combinations in order to make inferences about the relationship between these two masking situations.

Experiment I

Since masking depends on the temporal relationship between two succeeding stimuli, the central variable in the investigation of this phenomenon has been the duration of the interval between the two stimuli, here referred to as the interstimulus interval (ISI). The extent of "back-

[1] This research was supported by Mental Health Postdoctoral Fellowship FPD-11, 589 and National Institutes of Health Training Grant 3TIGM-1064. The author is indebted to Hans-Lukas Teuber for his helpful suggestions.

ward" masking is a complex function of the duration of the ISI. In some cases masking declines steadily with increasing ISI; in others it increases at first and then declines (Alpern, 1953; Kolers, 1962). Furthermore, masking studies have reported a wide range of ISIs at which maximal interference is observed. The underlying variables have not been clearly established, although some mechanisms have been proposed (Kolers, 1962).

This experiment investigated monoptic and dichoptic visual masking with a pattern or a flash as masking stimuli. The extent of masking was determined for different durations of ISI.

Method

Subjects.—The *S*s were four paid volunteers with equal visual acuity for both eyes (20/20 for all *S*s). These were highly trained *S*s who had participated in several other experiments on visual masking. Trained *S*s were used in order to minimize practice effects (Schiller & Wiener, 1963).

Apparatus.—The apparatus was a five-field tachistoscopic viewing box similar to the one described in a previous study (Schiller & Wiener, 1963). The electronic programmer was built by Sky Instrument Company. The light sources were cold-cathode mercury-argon lamps coated with magnesium-tungstate phosphor. For a more detailed description of programmer and lamps, see Kolers (1962).

Materials.—Test stimuli: The stimuli to be detected were letters of the alphabet (F, I, J, M, Q, and W were omitted). These letters were contact photographs (Matte finish) of No. 118, 24-pt. Future Medium "Lettraset" letters. Each letter appeared in the center of a 5 × 5 in. white card which was mounted on a plastic (Bakelite) plaque used for insertion in the viewing box.

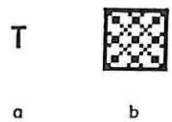


FIG. 1. The stimuli. (a) Sample of a test letter. (b) The pattern masking stimulus in Exp. I, II, and III.

Masking stimuli: Two masking stimuli were used: (*a*) a pattern (see Fig. 1) the area of which was 50% white and 50% black; (*b*) a bright flash of light measuring 3.5 foot-candles (ftc.) at the eyepiece. This light was obtained by increasing the luminance of the light sources in the proper field. This light was reflected from a blank white card mounted on a stimulus plaque. The usual illumination of each of the fields with blank cards as stimuli was .14 ftc. at the eyepiece. The reflectance of the white surfaces used was 90%; the black surfaces had a reflectance of 9%.

The *S*s looking into the viewing box saw a white circular surface (visual angle 7°6′). All stimuli were presented foveally, i.e., at the fixation point designated by a small red cross (visual angle 3′) in the center of the field. The visual angle of the letters was 30′; the pattern subtended an angle of 1°12′, and the flash an angle of 7°6′. A larger visual angle was used for the flash in order to minimize possible figural effects due to closeness of borders between the masked and masking stimuli.

Procedure.—Recognition thresholds were obtained for the singly presented test stimuli (letters) when followed by (*a*) no masking (N), (*b*) a flash to the same eye (FM), (*c*) a flash to the opposite eye (FD), (*d*) a pattern to the same eye (PM), and (*e*) a pattern to the opposite eye (PD). Four different ISIs were employed: 3, 20, 40, and 60 msec. For each of four *S*s, 16 threshold measurements were obtained for each of the ISI and masking conditions. The *S*s participated in four testing sessions. Each session began by taking two threshold measurements for warm-up using numbers as test stimuli but not followed by masking stimuli. During the rest of each session, four successive threshold measurements were taken (using letters) for each of the conditions. The order of test stimuli and conditions was random.

The masking stimuli, which always appeared *after* the test stimuli, were exposed for 100 msec. under all conditions. Recognition thresholds for the test stimuli were obtained by the method of ascending limits and were taken in 2-msec. increments. The ISI between the two stimuli refers to the interval between the offset of the test stimulus and the onset of the masking stimulus. During the ISI, the focus field came on again.

The interval between presentations was approximately 3 sec. The *E* alerted *S* by saying "ready" before each exposure. All *S*s were asked to relax briefly while conditions were changed, and *S*s were instructed to make

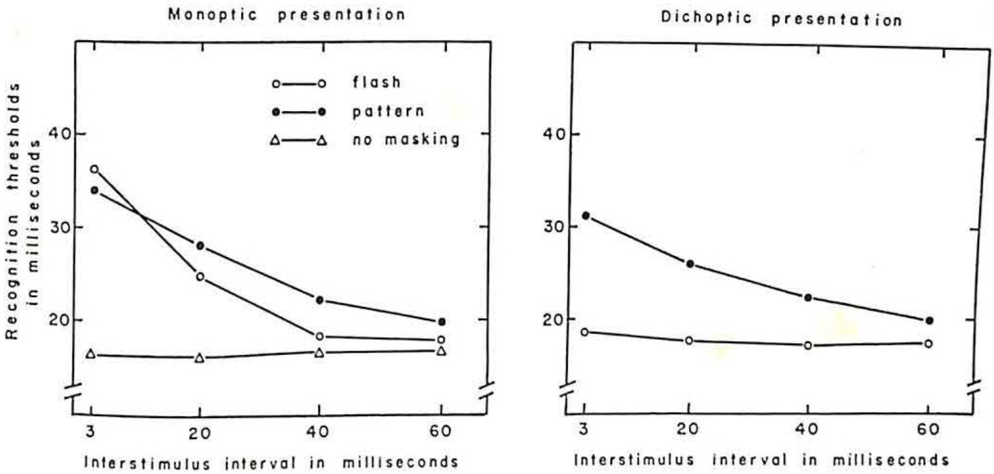


FIG. 2. Mean recognition thresholds for the conditions of Exp. I.

their best guesses. Each session lasted approximately 1 hr.

The eye not receiving the masking stimulus was exposed to a blank field under the same illumination as that of the focus field. Thus the illumination to the eyes remained constant except for the appearance of a letter and/or flash or pattern.

Results

The recognition threshold was the lowest exposure duration of the test stimulus at which *S*s correctly identified a letter. Since *S*s showed the same trends, only the combined data are presented here.

The mean recognition thresholds for the conditions of this study are presented in Fig. 2. The *SD*s (not shown) averaged 1–3 msec. for the N, FM, and FD conditions, and 2–6 msec. for the PM and PD conditions.

The results show that masking with the pattern occurred both monoptically and dichoptically. The extent of this masking was slightly less dichoptically. This difference for the mean thresholds, although consistent, was not significant for the four *S*s on t tests (t = 1.76, 1.41, 1.60, 1.31).

With the flash, masking occurred primarily on monoptic presentation. A slight masking effect was also found dichoptically with an ISI of 3 msec. The differences between DF and N conditions at an ISI of 3 msec. were significant for three of the four *S*s (t = 5.17, 3.30, 2.08, 1.53).

The extent of monoptic masking by flash or pattern (FM or PM) decreased with increasing ISI. The rate of decrease, however, was greater for Cond. FM than PM. An analysis of variance comparing FM and PM at the four ISI durations showed significant interaction between ISI and the masking stimulus employed ($F = 17.22$, $p < .01$).

EXPERIMENT II

The aim of the second experiment was to explore further the differences observed in visual masking when either a pattern or a flash was used as masking stimuli. In an earlier study (Schiller & Wiener, 1963) there was evidence that with a pattern as a masking stimulus there were decrements in recognition threshold with practice. Is this practice effect common to all masking situations or does it occur only with certain kinds of stimulus conditions?

Method

Subjects.—The *S*s were eight paid volunteers with equal visual acuity in both eyes. These were naive *S*s who had not previously participated in a visual masking experiment. The *S*s were divided in two groups.

Procedure.—The same apparatus and materials as in Exp. I were employed. All *S*s participated in eight testing sessions. During each of these sessions, 10 threshold measurements were taken for each of five conditions. These were the same as in Exp. I. no masking (N), monoptic flash (FM), dichoptic flash (FD), monoptic pattern (PM), and dichoptic pattern (PD). Five successive threshold measurements were taken for each condition, with each condition appearing twice during a session. Orders of conditions and test stimuli were random. The general procedure was identical to the one used in Exp. I.

For Group I, ISI was constant at 3 msec. For Group II, the ISI was always 20 msec.

Results

Since all the *S*s showed the same trends, only the combined data are presented here. Figures 3 and 4 show the mean recognition thresholds for the two groups for the 8 days of testing. With an ISI of 3 msec., the *SD*s for the PD and PM conditions averaged 4–9 msec., and for the FD, FM, and N conditions, 2–7 msec. With an ISI of 20 msec., the *SD*s for the PD and PM conditions averaged 2–6 msec., and for the FD, FM, and N conditions, 2–4 msec. The *SD*s in this experiment are higher than in Exp. I; this is probably due to the fact that naive *S*s participated in this experiment.

The results show that decrement in threshold with the pattern was much greater than the decrement with the flash. Table 1 shows the mean improvement in percent from Days 1 to 8 for all conditions.

The extent of masking by the pattern was somewhat greater monoptically than dichoptically. This difference was significant for seven of the eight *S*s according to *t* tests (t = 8.40, 6.74, 6.50, 3.71, 3.59, 3.46, 2.66, .85) for the overall means.

The extent of masking by the pattern was less with an ISI of 20 msec. than with an ISI of 3 msec. This difference was much greater for the monoptically presented flash. This finding confirms those of Exp. I with respect to the greater rate of decre-

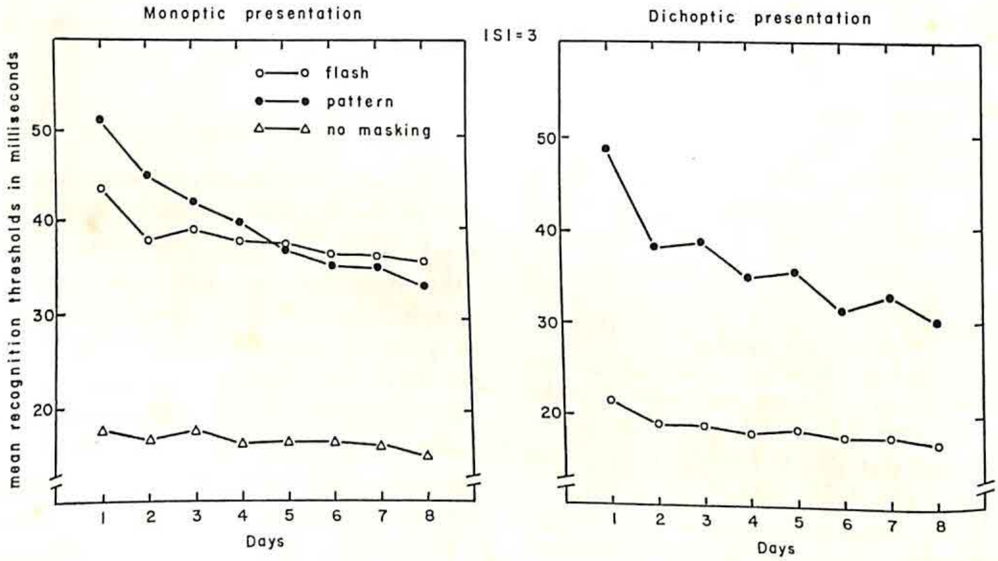


FIG. 3. Mean recognition thresholds for the eight sessions with an ISI of 3 msec.: Exp. II.

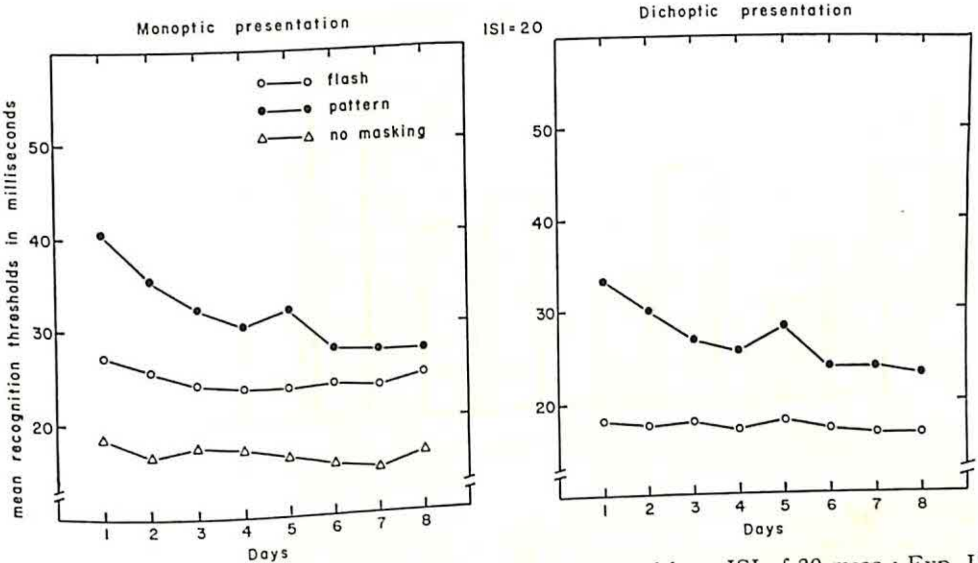


FIG. 4. Mean recognition thresholds for the eight sessions with an ISI of 20 msec.: Exp. II.

ment of masking by flashes as ISI is increased.

EXPERIMENT III

In the third experiment various combinations of flash and pattern were employed in the masking situation in a further attempt to assess the processes underlying masking and the relationship between the two kinds of masking situations.

Method

Subjects.—Two *S*s served in this study; both were highly trained.

Procedure.—The same apparatus and materials were used. The *S*s participated in four testing sessions. A constant ISI of 3 msec.

TABLE 1

PERCENT IMPROVEMENT OF RECOGNITION THRESHOLDS (IN MILLISECONDS) IN EXP. II FROM DAY 1 TO DAY 8

	ISI 3	ISI 20
N	15.5	15.3
FM	17.5	8.4
FD	20.4	8.9
PM	34.3	32.5
PD	37.9	30.9

was employed. There were nine conditions in this study. The first five were similar to the conditions used in Exp. I and II. (Cond. 1) N, (Cond. 2) FM, (Cond. 3) FD, (Cond. 4) PM, and (Cond. 5) PD. Conditions 6–9 employed various pairings of flash and pattern as follows: (Cond 6) FPM. The flash and pattern appeared together; both were presented to the same eye as the test stimulus. (Cond. 7) FPD. The flash and pattern appeared together; both were presented opposite to the eye which received the test stimulus. (Cond. 8) PD-FM. The flash was presented to the same eye as the test stimulus; at the same time, the pattern was exposed to the other eye. (Cond. 9) PM-FD. The pattern was presented to the same eye as the test stimulus; at the same time, the flash was exposed to the other eye.

Presenting the flash and pattern together (FPM and FPD) had the effect of increasing the contrast of the pattern, since the more intense light sources were now reflected not from a blank card as in Cond. FM and FD but from the card containing the pattern.

Results

The results of this experiment are reported in Fig. 5. Since both *S*s showed the same trends, their data were combined. The differences among conditions were analyzed by *t* tests. All comparisons yielded sig-

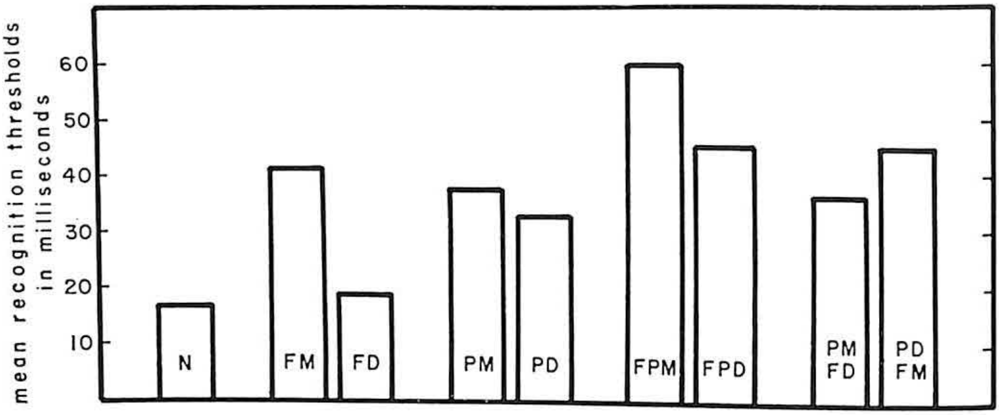


FIG. 5. Mean recognition thresholds for the conditions of Exp. III with an ISI of 3 msec.

nificant differences at the .01 level for both *S*s except for the following: FM and PM, FM and PM-FD, PM and PM-FD. The difference between N and FD was significant for only one of the *S*s. Conditions 1–5 produced results similar to those found in Exp. I and II. These were essentially controls to be compared with Cond. 6–9. Presenting flash and pattern together to one eye (Cond. FPM and FPD) significantly increased the masking effect. This occurred upon monoptic and dichoptic presentation, but masking on monoptic presentation was significantly greater than on dichoptic.

If the flash interferes with perception of the test stimulus at a stage prior to that at which the two monocular fields are combined, it would be reasonable to assume that under the FM-PD condition the perceiver has to overcome first the interference due to the flash given to the same eye, and then the interference due to the pattern given to the other eye. Accordingly, a higher recognition threshold should be obtained with the FM-PD condition than with the FM and PD conditions. The results bear out this expectation.

Discussion

The results of this study show the following:

1. Backward masking by a flash of light is primarily a monoptic effect; masking by pattern occurs under monoptic and dichoptic conditions.

2. Both masking by pattern and masking by monoptically presented flash decrease with increasing ISI. The rate of this decrease is greater for the flash.

3. Repeated trials with patterns as masking stimuli lead to a decrement in the masking effect. Such "practice effects" are not obtained for flashes as masking stimuli.

4. Pairing the flash and the pattern as masking stimuli produces an increase in the masking effect under both monoptic and dichoptic conditions as compared with effects of either flash or of pattern presented alone.

5. Presenting the flash to one eye and the pattern to the other increases their combined masking effect, provided the flash is given under monoptic conditions but not if the flash is given under dichoptic conditions.

These results suggest that there are two distinguishable processes of masking involved in this study. One of these (for flash) seems to occur in the central visual pathways prior to the mixing of the two monocular representations. The other (for pattern) occurs at levels where the

two monocular fields interact; the decline in masking by pattern on repeated trials further suggests complex processes of perception in which learning might play a role. The type of effect observed with the flash seems to be related to the interference effects reported by Boynton and Kandel (1957), who were apparently the first to introduce the term "masking" for this phenomenon.

The visual masking effects reported here seem to be different from the interference which takes place by metacontrast (Alpern, 1953). In metacontrast interference with the first stimulus is minimal at short ISIs, reaches a maximum between 75–100 msec., and then declines again. In the masking experiments reported here, interference declined steadily with increasing ISI.

The results reported here show a small degree of masking with a dichoptic flash. The extent of this masking varied to some degree with *S*s, and two practiced *S*s did not show a significant effect at all. On the basis of such variability it is difficult to draw any conclusions from these findings. It should be noted, however, that small but reliable dichoptic effects, using flashes of light for both the masked and masking stimuli, have been reported by Battersby and Wagman (1962) and Kandel (1958).

The extent of masking by pattern was slightly less dichoptically than monoptically. This difference can apparently be magnified when the monoptic and dichoptic conditions, rather than being randomized as in this study, are presented in separate series of trials (Schiller & Wiener, 1963). It is possible that when these conditions are not randomized the eye receiving the pattern under dichoptic conditions becomes slightly inhibited with practice, thereby decreasing the extent of masking.

REFERENCES

Alpern, M. Metacontrast. *J. Opt. Soc. Amer.*, 1953, **43**, 648–657.

Battersby, W. S., & Wagman, I. H. Neural limitations of visual excitability. IV. Spatial determinants of retrochiasmal interaction. *Amer. J. Physiol.*, 1962, **203**, 359–365.

Boynton, R. M. Some temporal factors in vision. In W. A. Rosenblith (Ed.), *Sensory communication*. New York: Wiley, 1961. Pp. 739–756.

Boynton, R. M., & Kandel, G. L. On responses in the human visual system as a function of adaptation level. *J. Opt. Soc. Amer.*, 1957, **47**, 275–286.

Exner, S. Über die zu einer Gesichtswahrnehmung nöthige Zeit. *S. B. Akad. Wiss. Wien*, 1868, **58**, 601–632.

Kandel, G. L. A psychophysical study of some monocular and binocular factors in early adaptation. Unpublished doctoral dissertation, University of Rochester, 1958.

Kolers, P. A. Intensity and contour effects in visual masking. *Vis. Res.*, 1962, **2**, 277–294.

Kolers, P. A., & Rosner, B. S. On visual masking (metacontrast): Dichoptic observation. *Amer. J. Psychol.*, 1960, **73**, 2–21.

Lindsley, D. B. Electrophysiology of the visual system and its relation to perceptual phenomena. In M. A. B. Brazier (Ed.), *Brain and behavior*. Vol. 1. Washington, D. C.: American Institute of Biological Sciences, 1961. Pp. 359–392.

Raab, D. H. Backward masking. *Psychol. Bull.*, 1963, **60**, 118–129.

Schiller, P. H., & Wiener, M. Monoptic and dichoptic visual masking. *J. exp. Psychol.*, 1963, **66**, 386–393.

Sperling, G. What visual masking can tell us about temporal factors in perception. In, *Proceedings of the Seventeenth International Congress of Psychology*. Amsterdam: North-Holland, 1964. Pp. 199–200.

Stigler, R. Chronophotische studien über den Umgebungskontrast. *Pflüg. Arch. ges. Physiol.*, 1910, **134**, 365–435.

Werner, H. Studies on contour: I. Qualitative analysis. *Amer. J. Psychol.*, 1935, **47**, 40–64.

Werner, H. Studies on contour strobostereoscopic phenomena. *Amer. J. Psychol.*, 1940, **53**, 418–422.

(Received January 13, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 200-206

APPLICATION OF A MARKOV MODEL TO FREE RECALL AND RECOGNITION [1]

WALTER KINTSCH AND CHARLES J. MORRIS
University of Missouri

In 2 experiments lists of nonsense syllables were learned by the methods of recognition and recall. In a 3rd experiment Ss first learned to recognize a list of nonsense syllables and then to recall it. The recognition data could be described by a simple 2-state Markov model. A 3-state Markov model was needed for the description of the recall data. It was hypothesized that recall learning involved 2 stages, a recognition stage and a 2nd stage where the response becomes available in the absence of the stimulus, and that each of these stages can be described as a simple Markov process. In Experiment III it was shown that recall learning after recognition learning can actually be described by a 2-state Markov model.

The results of many studies conducted with the method of free recall and recognition permit the generalization that recognition is a simpler process than recall. The low incidence of extralist intrusions in studies of free recall provides indirect evidence that recognition of list membership develops faster than recall (Deese, 1959). When recognition and recall are employed as measures of retention the former yields higher scores than tests of recall (Postman & Rau, 1957). Apparently an *S* first learns to recognize an item and then he learns to recall it. Thus, learning by the method of free recall can be regarded as a two-stage process. If one is willing to assume that learning in each stage is all-or-none, a simple model can be obtained for this two-stage process. This assumption appears justifiable because learning in each stage is elementary, well-structured, and well-controlled and because the application of Markov models to such learning tasks has been notably successful (Bower, 1961, 1962; Bower & Trabasso, 1963; Estes, 1960, 1961; Suppes & Ginsberg, 1963).

[1] This study was supported by Research Grant GB-195 from the National Science Foundation.

The strategy to be followed here consists in analyzing free recall into two one-stage learning processes. The model used to describe the elementary learning process is a simple two-state Markov model of the type used by Bower (1961). This model assumes a constant probability on every learning trial that an item will move from the starting state to the learning state. The learning state of the first subprocess will be the starting state for the second subprocess. The probability of a correct response in each state will depend both upon *E*'s procedure and the response biases of *S*. In general, subprocesses will have different learning rates.

Using the conventional notation for Markov processes the model for recognition learning is given by

$$\begin{array}{c} \\ C_0 \\ C_1 \end{array} \begin{array}{c} \begin{array}{cc} C_0 & C_1 \end{array} \\ \begin{bmatrix} (1-c) & c \\ 0 & 1 \end{bmatrix} \end{array} \quad \begin{array}{c} \text{Probability of a correct response} \\ 1-p \\ 1 \end{array} \quad . \qquad [1]$$

This model is identical with Bower's paired-associate model, except for the trivial modification that the process always starts in State C_0 in Bower's case, whereas in the typical recogni-

tion experiment data are collected only after one learning trial has been given. Thus the expressions given by Bower (1961) must be adjusted slightly. They are reproduced in the Appendix.

Free recall will be represented by a three-state Markov process. The subprocess of recognition is again described by Matrix 1. However, the probabilities of a correct response in C_0 and C_1 will be zero and $(1 - r)$ in this case. State C_1 is the learning state for the recognition subprocess and the initial state for the recall subprocess. In this state an item can be recognized, but active reproduction is not always possible. The recall process itself is also described by a transition matrix like Matrix 1, which gives as a total process

$$
\begin{array}{c} \\ C_0 \\ C_1 \\ C_2 \end{array}
\begin{array}{c} \begin{array}{ccc} C_0 & C_1 & C_2 \end{array} \\ \left[\begin{array}{ccc} (1-c) & c & 0 \\ 0 & (1-\theta) & \theta \\ 0 & 0 & 1 \end{array}\right] \end{array}
\quad
\begin{array}{c} \text{Probability of a correct response} \\ 0 \\ 1-r \\ 1 \end{array} . \qquad [2]
$$

This three-state model is similar to the one used by Bower and Theios (1964) and Kintsch (1963), except that allowance was made there for the possibility that the learning rate in the second stage might be different after a failure than after a success. This refinement was abandoned here for the sake of simplicity.

Theoretical predictions from Matrix 2 are given in the Appendix. They are identical with the ones given by Bower and Theios (1964) except for the modification noted above. When these expressions are compared with the experimental results, a "first" trial on which all items are missed will be included in the data.

Experiments I and II provide data for a test of the recognition and recall model. In Exp. III the learning process is experimentally subdivided into a recognition and a recall phase.

Method

Subjects.—Fifty students from an introductory psychology course served as *S*s, 16 each in Exp. I and II, and 20 in Exp. III. The *S*s were randomly assigned to the three experimental groups.

Materials.—Sixty nonsense syllables were selected from Noble's (1961) list. The association value of all syllables was between .5 and .6. Each syllable was printed on a 3 × 5 in. index card.

Procedure

Experiment I: Recognition.—Each *S* received a list of 15 nonsense syllables which were selected randomly from the pool described above. Each syllable was presented to *S* for 2 sec. After all syllables had been presented the cards were shuffled together with 15 syllables from the same pool which had not been shown before and were presented again. The *S* was now instructed to respond with "yes" or "no" to each card, depending upon whether he thought he had seen it before or not. This procedure was repeated until *S* reached a criterion of three successive correct recognition trials. All *S*s were instructed to spell out the syllables aloud as they were shown. The *E* recorded for each of the 15 original items whether it had been recognized.

Experiment II: Recall.—The *S*s were given 10-item lists to learn at the rate of approximately one item every 2 sec. After each presentation *S* was given up to 120 sec. during which he attempted to recall the items verbally. Again, a criterion of three successive correct trials was required.

Experiment III: Recognition followed by recall.—Ten-item lists were employed in this experiment. First the list was learned to a criterion of three successive correct trials with the procedure of Exp. I. The *S*s were then instructed to try to recall as many items as they could. Further trials were given, following the procedure of Exp. II, until a criterion of three successive correct recalls was reached.

Results

Analysis of learning stages.[2]—According to the model given by Ma-

[2] Failures to recall or recognize will be referred to as errors.

trices 1 and 2 different stages of learning are characterized by constant response probabilities. In Exp. I the probability of a correct response should not change as long as an item is in State C_0. Since an item is in the starting state at least until the trial of the last error, response probability should be constant on trials before the last error. To test this proposition sequences before the last error were divided into thirds and the errors in each third were recorded. They were 8, 6, and 9, respectively, out of a total number of 46 observations in each third. A chi-square test for the stationarity of this Vincent curve yielded a value of .786 which was obviously not significant.

In Exp. II stationarity was expected in the intermediary state according to Matrix 2. Again it was impossible to know on exactly which trials the process was in the intermediary state, but all trials between the first success and the last error must belong to this state. Thus Vincentized sequences over these trials were examined for stationarity. Halves were used instead of thirds this time because not enough long sequences were available. The number of errors in the two halves were 21 and 20 out of 59 observations, with a corresponding nonsignificant chi square of .37. This aspect of the recall data is therefore in agreement with the hypothesis that recall is a two-stage process. To show that recall could not be described as a one-stage process, a Vincent curve was constructed in the manner of Exp. I. The hypothesis of stationarity had to be rejected for this curve. The proportion of errors was .61 and .43 in the two halves, $\chi^2 = 9.349$, $p < .01$.

In Exp. III the two stages of recall were separated experimentally. Performance before the last error should therefore be stationary in each phase, as it was in Exp. I. Learning in the recognition phase was, however, so fast that no meaningful tests of stationarity could be performed. In the recall phase the numbers of errors on trials previous to the last error were 17 and 20 for the first and second halves of these trials (out of 33 observations in each half). The number of errors thus actually increased from the first to the second half. The difference was, however, not statistically significant ($\chi^2 = .553$).

The stationarity data are presented here as evidence that a Markov model can be used to describe recognition and recall learning. They are not intended to prove that learning was all-or-none or that no other interpretations are feasible, but merely to show that the very simple conceptualization in terms of Markov processes is in agreement with this aspect of the data. Another requirement of the Markov model, namely that trials before the last error be independent, was also met by the data. Chi-square tests of independence of successive trials yielded a value of 2.66 for Exp. I and .87 for the intermediary trials in Exp. II, both with $df = 1$. In Exp. III a chi square of .44 was obtained for the recall phase ($df = 1$), while not enough data were available for a statistical test in the recognition phase.

Fit of the Model

Experiment I.—Further comparisons of the model with the recognition data necessitated the estimation of the two parameters of the model. An intuitively reasonable estimate for p was available in the proportion of errors on trials before the last error, $\hat{p} = .214$. An estimate for the conditioning parameter was obtained by setting the theoretical expression for the total number of errors equal to the

observed total number of errors, substituting for p and solving for c. The value obtained in this manner was $\hat{c} = .282$. The predictions from the model were obtained by a simple modification of the expressions given by Bower (1961). These are included in the Appendix. Using the estimates p and c, numerical predictions for various statistics of the data were then calculated. In general the fit between these predictions and the data was satisfactory. A few comparisons are given in Table 1.

Experiment II.—In order to derive predictions from the recall model, the estimation of three parameters was

TABLE 1

Predictions and Observed Statistics in Exp. I

	Observed	Predicted
Mean number of errors	.55	—
SD	.89	.92
Mean trials before the first recognition	.25	.18
SD	.53	.46
Mean trial of last error	1.17	1.23
SD	2.14	2.45
Total number of error runs	.46	.46

necessary. The proportion of errors on trials between the first recall and the last error served as an estimate for r, $\hat{r} = .333$. From Matrix 2 an expression was derived for the number of times a success is followed by an error on the next trial. This expression depends only upon θ and upon r. By substituting $\hat{r}$ into the theoretical expression and setting it equal to its observed value, an equation was obtained and solved for θ. The obtained estimate was $\hat{\theta} = .297$. The estimate for c was then obtained as in Exp. I, using $\hat{p}$ and $\hat{\theta}$. It was found to be $\hat{c} = .502$. These estimates were used to obtain the predictions which are compared with the data in Table 2.

TABLE 2

Predictions and Observed Statistics in Exp. II

	Observed	Predicted
Mean number of errors	3.11	—
SD	1.73	1.91
Mean number of trials before the first recall	2.40	2.42
SD	1.45	1.58
Mean trial of last error	4.14	4.10
SD	2.68	3.09
Total number of error runs	1.52	1.52

Experiment III.—Phase I of this experiment was analyzed exactly like the recognition data in Exp. I. The estimation procedure yielded $\hat{p} = .178$ and $\hat{c} = .372$. Table 3 provides a comparison between observed and predicted statistics from which the goodness of fit of the model can be judged.

The one-stage model applied to the recall phase of this experiment was somewhat different from the model used to describe recognition data. At the end of recognition learning several items were already available for recall, in spite of the fact that *S* was not instructed to recall items or was previously tested for recall. The proportion of such items (λ) was therefore another parameter of the one-stage model. The theoretical expressions derived from this modified model are

TABLE 3

Predictions and Observed Statistics in Recognition Phase in Exp. III

	Observed	Predicted
Mean number of errors	.30	—
SD	.56	.62
Mean number of trials before the first recognition	.17	.13
SD	.44	.37
Mean trial of last error	.54	.62
SD	1.12	1.52
Total number of error runs	.23	.27

TABLE 4

Predictions and Observed Statistics: Recall Phase of Exp. III

	Observed	Predicted
Mean number or errors	.62	—
SD	1.00	1.01
Mean number of trials before the first recall	.42	—
SD	.80	.78
Mean trial of last error	.83	.81
SD	1.45	1.39
Total number of error runs	.42	.43

also included in the Appendix. A simple estimate for the proportion λ and the conditioning parameter c was obtained by solving the two equations for the total number of errors and the number of errors before the first recall for c and λ. In order to do this an estimate for p was first obtained in the usual manner. The resulting parameter estimates were $\hat{p} = .577$, $\hat{\lambda} = .487$, and $\hat{c} = .474$. A comparison of the predictions obtained from this model with the data of the recall phase is presented in Table 4. It is interesting to note that the number of items available for recall at the beginning of Phase II was approximately equal to those learned in one regular recall trial.

In applications of the model it was assumed that all items and all *S*s were equivalent. Experiment III provides a crude test of this assumption, at least as far as the equivalence of *S*s is concerned. If individual differences were significant in this experiment slow learners in Phase I should also be slow learners in Phase II. According to the model, on the other hand, slow and fast learning of an item was regarded as a random process governed by one and the same conditioning rate. The correlation between number of trials to criterion in Phases I and II should therefore be zero. A rank-order correlation of .261 was actually obtained, which was not significantly different from zero.

Discussion

The results of this experiment support the hypothesis that learning by the method of recall is a two-stage process, with learning to recognize as a first stage. Secondly, the stationarity data reported and the successful fit of the model demonstrate that it is possible to describe each stage as a two-state Markov process. Concerning this second point some experience from other experiments and some speculations must be reported. It would be quite easy to destroy the all-or-none character of the data. This can be achieved by using learning material which is so complex that even recognition learning involves more than one step. Low-association-value consonant trigrams will do for that purpose. Also, the application of the present model is precluded whenever *E* unbalances the equivalence of the items to be learned. This can be done by allowing some items to be very different from others, by presenting the items always in the same order, or by any other means which ensure that some items will arouse more attention than others. What happens then is that certain items are set aside for a few trials while *S* concentrates, for example, on the first and last items of the list. Increase in list length might have the same effect. With long lists *S*s apparently do not divide their attention evenly over all items, but single out some for learning and leave the others for later. A one-stage model obviously would be inadequate to describe the learning of the latter items.

The question of whether the recall learning in Exp. II or the recognition and recall learning in Exp. III was more efficient can not be answered unequivocally. In terms of total number of errors made, the procedure used in Exp. III was much superior. However, such a direct comparison is inadmissible because the experimental operations were such as to ensure errors in the first stage in Exp. II but not during the same stage in Exp. III. A comparison in terms of the total number of trials to complete learning in Exp. II and III revealed no difference. The mean number of trials to criterion was 9.4 and 11.1 in the two

experiments, $t\ (34) = 1.90$, $p > .05$. It should be noted, however, that the score for Exp. III includes a total of six criterion trials (three successive correct trials in both phases of the experiment). Only three criterion trials are included in the score for Exp. II.

The two stages which are postulated by the present model are recognition learning and recall. Recognition learning consists in differentiating the items of a list, i.e., of encoding the stimulus. This encoding process establishes a response to an item when it is presented. In the beginning of the second phase the response is thus elicited by the sensory stimulus, but it is not reliably available in the absence of the sensory stimulus. The second learning phase consists in making the response always available. This interpretation is in agreement with the different functions of recall and recognition as described by Postman and Rau (1957). It also has certain features in common with a stochastic model for free recall described by Waugh and Smith (1962). Corresponding to the two stages which were identified here, Waugh distinguishes between processes of labeling and fixation. Her model is also a Markov model, but somewhat more complex than the present model. The most important difference lies in the fact that Waugh allows for direct transitions from the starting state to the learning state. This is incompatible with the idea of analyzing recall into a sequence of two one-step processes.

The model for free recall by Miller and McGill (1952), (also Bush & Mosteller, 1955) differs from the present model in interpretation as well as formal structure. The two models do, however, lead to some common predictions. For instance, they predict a geometric or nearly geometric distribution of trials before the first recall (see also Murdock & Babick, 1961).

Although the present model accounts for some aspects of free-recall data, it does not give a complete description. The persistent clustering of responses in free-recall data is the most significant aspect of the data which has been neglected.

REFERENCES

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255–280.

Bower, G. H. A model of response and training variables in paired-associate learning. *Psychol. Rev.*, 1962, **69**, 34–53.

Bower, G. H., & Theios, J. A learning model for discrete performance levels. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1964.

Bower, G. H., & Trabasso, T. Reversals prior to solution in concept identification. *J. exp. Psychol.*, 1963, **66**, 409–418.

Bush, R. R., & Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

Estes, W. K. Learning theory and the new mental chemistry. *Psychol. Rev.*, 1960, **67**, 207–223.

Estes, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika*, 1961, **26**, 73–84.

Kintsch, W. All-or-none learning and the role of repetition in paired-associate learning. *Science*, 1963, **140**, 310–312.

Miller, G. A., & McGill, W. F. A statistical description of verbal learning. *Psychometrika*, 1952, **17**, 361–396.

Murdock, B. B., Jr., & Babick, A. J. The effect of repetition on the retention of individual words. *Amer. J. Psychol.*, 1961, **74**, 596–601.

Noble, C. E. Measurement of association value (a), rated associations (a') and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychol. Rep.*, 1961, **8**, 487–521.

Postman, L., & Rau, L. Retention as a function of the method of measurement. *U. Calif. Publ. Psychol.*, 1957, **8**, 217–270.

Suppes, P., & Ginsberg, R. A fundamental property of all-or-none models. *Psychol. Rev.*, 1963, **70**, 139–161.

Waugh, N. C., & Smith, J. E. K. A stochastic model for free recall. *Psychometrika*, 1962, **27**, 141–154.

(Received January 15, 1964)

APPENDIX

Theoretical Predictions

A. Some derivations from the two-state model when observations are taken only after the first learning trial:

Distribution of number of errors per item, T:

$$E(T) = \frac{(1-c)p}{c}$$

$$\mathrm{Var}(T) = (E(T))^2 + E(T)$$

Distribution of number of trials before the first success, J:

$$E(J) = \frac{p(1-c)}{1-p(1-c)}$$

$$\mathrm{Var}(J) = (E(J))^2 + E(J)$$

Distribution of trial of last error, N:

$$E(N) = \frac{p(1-c)}{c(1-(1-c)(1-p))}$$

$$\mathrm{Var}(N) = E(N)\frac{2-c}{c} - (E(N))^2$$

Total number of error runs, R:

$$E(R) = E(T)(1-(1-c)p).$$

B. Derivations from the same model when already a proportion λ of the items are in the learning state on Trial 1:

Distribution of total number of errors per item, T:

$$E(T) = \frac{p(1-\lambda)}{c}$$

$$\mathrm{Var}(T) = E(T)\frac{2-b}{b} - (E(T))^2$$

$$\text{where } b = \frac{c}{1-(1-c)(1-p)}$$

Distribution of number of trials before the first success, J:

$$E(J) = \frac{p(1-\lambda)}{1-p(1-c)}$$

$$\mathrm{Var}(J) = E(J)\left\{\frac{2p(1-c)}{1-p(1-c)} + 1\right\} - (E(J))^2$$

Distribution of trial of last error, N:

$$E(N) = \frac{p(1-\lambda)}{c\{1-(1-c)(1-p)\}}$$

$$\mathrm{Var}(N) = E(N)\frac{2-c}{c} - (E(N))^2$$

Total number of error runs, R:

$$E(R) = E(T)(1-(1-c)p).$$

C. Derivations from the three-state model:

Distribution of the total number of errors per item, T:

$$E(T) = \frac{1}{c} + \frac{r}{\theta}$$

$$\mathrm{Var}(T) = (1-r)b\frac{2-c}{c^2} + \frac{cb(1-(1-r)b)}{c-b} \times \left\{\frac{2-b}{b^3} - \frac{2-c}{c^3}\right\} - (E(T))^2$$

$$\text{where } b = \frac{\theta}{1-(1-\theta)(1-r)}$$

Distribution of number of trials before the first success, J:

$$E(J) = \frac{1}{c} + \frac{r}{1-r+r\theta}$$

$$\mathrm{Var}(J) = (1-r)\frac{2-c}{c^2} + \frac{cr(1-a)}{a-(1-c)} \times \left\{\frac{1+a}{(1-a)^3} - \frac{2-c}{c^3}\right\} - (E(J))^2$$

$$\text{where } a = r(1-\theta)$$

Distribution of trial of last error, N:

$$E(N) = \frac{1}{c} + \frac{r}{\theta(r+(1-r)\theta)}$$

$$\mathrm{Var}(N) = (1-r)b\frac{2-c}{c^2} + \frac{brc}{\theta-c} \times \left\{\frac{2-c}{c^3} - \frac{2-\theta}{\theta^3}\right\} - (E(N))^2$$

Total number of error runs, R:

$$E(R) = \frac{r}{\theta}(1-(1-\theta)r) + (1-r)$$

Average number of times a success is followed by an error one trial later, d:

$$E(d) = \frac{r(1-r)(1-\theta)}{\theta}.$$

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 207–208

SUPPLEMENTARY REPORT

"FATE" OF LIST 1 R-S ASSOCIATIONS IN TRANSFER THEORY [1]

NORMA R. ELLINGTON AND DONALD H. KAUSLER
St. Louis University

Extinction of List 1 R-S associations during practice on List 2 was tested for the A-B, C-B transfer paradigm by means of the Barnes and Underwood (1959) modified recall method. It was found that both the R-S associations of List 1 and the stimuli themselves became increasingly unavailable for recall with increasing practice on List 2. The R-S associations and stimuli of List 2, however, became increasingly available for recall with practice on List 2. A control group, without List 2 practice, indicated that the decrement in R-S recall on List 1 could not be attributed to simple forgetting occurring during List 2 practice.

Evidence for the extinction of List 1 A-B associations during the learning of List 2 A-C associations has been presented by both Barnes and Underwood (1959) and Goggin (1963). A similar phenomenon might be expected for the R-S associations of List 1 pairs in an A-B, C-B transfer paradigm in that the R-S associations of the two lists enter into an A-B, A-C relationship (Kausler & Kanoti, 1963; Twedt & Underwood, 1959). In fact, some evidence for the extinction of B-A (i.e., List 1 R-S associations) associations following practice on List 2 C-B (S-R) pairs has been presented by Keppel and Underwood (1962). The present study extended this area of investigation by replicating the Barnes and Underwood experiment on the C-B paradigm, with the extinction of B-A associations being measured by their modified free recall technique after varying degrees of List 2 practice.

Method.—Seventy-five *S*s (undergraduate students in psychology) were assigned alternately to four experimental groups and a control group. Each group practiced on List 1 (A-B pairs) to a criterion of one perfect trial. The four experimental groups received (following a familiarization trial) either 1, 5, 10, or 20 trials on List 2 (C-B pairs). The control group performed on a perceptual problem-solving task, in lieu of List 2, for a period of time approximating that spent by the 20-trial group on List 2. Pairs on both lists were presented at a 2:2-sec. rate, with a 4-sec. intertrial interval. Three different serial orders were employed for each list.

Following List 2 practice (or the problem-solving task), R-S recall for the associations of both lists was measured by presenting *S* with a blank sheet of paper and a second sheet, which was available throughout the R-S recall, containing the responses of the two lists in alphabetical order. The *S*s were instructed to write down all of the stimuli from the two lists they could remember and in whatever order they came to mind. No time limit was imposed for this task. They were then instructed to write beside each stimulus the response with which it had been paired and also to indicate whether the stimulus came from List 1 or 2 (again without a time limit). Following this, *S*s were told they could add as many other stimuli as they remembered at that time, even if they were not absolutely sure of all of the letters. The R-S recall data were therefore comparable to the Method 1 and Method 2 S-R data of Goggin's (1963) study.

The stimuli for the two lists were nonsense syllables selected from the 60 to 73% range of Archer's (1960) standardization, with both intra- and interlist similarity minimized as much as possible. The response terms were selected from the adjectives employed by Barnes and Underwood.

Results.—For all five groups combined, the mean number of trials to criterion on List 1 was 11.48. Comparability among the five groups in List 1 performance is attested to by the lack of a statistically significant difference between the means for trials to criterion, $F(4, 70) = 1.01$, $p > .25$. R-S recall scores were considered correct in Method 1 (after Goggin, 1963) if the stimulus term was paired with the right response term and given the correct list number. In Method 2 (after Goggin) the stimulus was counted correct if it was recalled, whether or not the response with which it was paired and the list number were remembered. Summary data for the four experimental groups on both lists and both methods are given in Table 1.

[1] This study is based on a thesis submitted by the first author to the Graduate School, St. Louis University in partial fulfillment for the Master of Arts degree.

TABLE 1

Summary Data for R-S Recall on Lists 1 and 2

Method	Trials on List 2: 1		5		10		20	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1								
List 1	4.13	1.64	1.93	1.62	2.07	2.19	1.40	1.12
List 2	2.13	1.92	4.33	2.77	4.27	1.67	6.80	1.47
2								
List 1	4.40	1.50	2.46	1.92	2.40	1.99	1.73	1.44
List 2	3.27	2.05	5.27	2.40	5.13	1.96	6.87	1.12

From Table 1 it is apparent that, with practice on List 2, not only are the R-S associations of List 1 extinguished, but also that the List 1 stimuli themselves become unavailable for recall. The analyses of variance for List 1 revealed a significant group effect on both recall scores—for Method 1, $F\ (3, 56) = 7.53$, $p < .001$; for Method 2, $F\ (3, 56) = 6.60$, $p < .001$. It is also apparent that, with increasing practice on List 2, both the stimuli and R-S associations of List 2 become increasingly available for recall. Analyses of variance again revealed a significant group effect on both recall scores—for Method 1, $F\ (3, 56) = 13.39$, $p < .001$; for Method 2, $F\ (3, 56) = 8.62$, $p < .001$.

The control group checked the possibility that the decline in List 1 R-S scores could be attributed to simple forgetting. The mean score on both methods for this group was 5.73, a value significantly higher than the 20-trial experimental group—t's (28) of 8.45 and 7.71 for Methods 1 and 2, respectively—indicating that the decrement in recall of B-A associations cannot be attributed to forgetting, but rather to the result of learning B-C associations.

Interpretation of the List 1 data for both Methods 1 and 2, however, is complicated by the presence of possibly confounding factors. Method 1, for example, may be measuring both degree of R-S extinction and the inability to differentiate between lists. That is, R-S associations required both recall *and* list identification in order to be scored as correct by this method. In an attempt to eliminate this differentiation factor, a corrected Method 1 score was computed for each *S* in the experimental groups by adding to his score all associations correctly recalled but placed in the wrong list. The lack of a pronounced list-differentiation effect, however, is indicated by the relatively slight change in means, with the corrected Method 1 means being 4.13, 2.00, 2.13, and 1.53 for the 1-, 5-, 10-, and 20-trial groups, respectively (see Table 1 for the uncorrected means). Moreover, increasing extinction with practice increments is still apparent from the statistically significant group effect, $F\ (3, 56) = 6.56$, $p < .001$.

Similarly, Method 2, as an index of stimulus availability, is limited by the procedure employed. To be scored as correct by this method, a stimulus had merely to be named, regardless of response pairing and list identification. The naming of stimuli under these conditions may be confined to those stimuli for which there are intact associations, unless *S* has other stimuli available for which he attempts to guess the response pairing and list membership. A guessing set was not intentionally generated in this study, and it is conceivable that Method 2 underestimates stimulus availability under these conditions. Consequently, it is difficult to estimate the degree of extinction of R-S associations beyond that involved in the unavailability of the stimuli per se. Nevertheless, it should be noted that for all four experimental groups the corrected Method 1 means are below the Method 2 means and that for two of the four groups the difference in means reached statistical significance—$p < .05$; correlated t's (14) = 1.80, 2.47, 2.29, and 1.39 for 1, 5, 10, and 20 trials, respectively.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, 74(10, Whole No. 497).

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, 58, 97–105.

Goggin, J. Supplementary report: Influence of the written recall measure on first-list associations. *J. exp. Psychol.*, 1963, 65 619–620.

Kausler, D. H., & Kanoti, G. A. R-S learning and negative transfer effects with a mixed list. *J. exp. Psychol.*, 1963, 65, 201–205.

Keppel, G., & Underwood, B. J. Retroactive inhibition of R-S associations. *J. exp. Psychol.*, 1962, 64, 400–404.

Twedt, H. M., & Underwood, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, 58, 111–116.

(Received January 4, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 2, 209–212

Beginning with the first issue of Volume 69 and continuing through the remainder of that volume and Volume 70 (1965), the titles and authors of accepted papers will be listed following Supplementary Reports. Such listing will allow readers to become aware of research many months in advance of journal publication. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing. This listing plus that published in the January 1965 issue are the entire backlog of manuscripts accepted by this journal. The articles listed below are scheduled to appear approximately 8 months hence.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

An Examination of the Two Stages of Paired-Associate Learning as a Function of Intralist Response Similarity (IRS) and Response Meaningfulness (*M*): John Jung*: 1726 Polk St., San Francisco, California.

Effects of Food and Water Deprivation on the Performance of a Response Motivated by Acquired Fear: Ronald Ley*: Graduate Faculty of Political and Social Science, New School for Social Research, 66 West 12th Street, New York 11, New York.

Sensitivity of Subjective Probability Revision: Cameron R. Peterson* and Alan J. Miller: Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan, P.O. Box 618, Ann Arbor, Michigan.

Positional Cues as Mediators in Discrimination Learning: Sheldon M. Ebenholtz*: Department of Psychology, Connecticut College, New London, Connecticut.

Perception of Deviations in Repetitive Patterns: Cord B. Sengstake*: Department of Physiology, School of Medicine, Center for Health Sciences, University of California, Los Angeles, California 90024.

Time Estimation and Increases in Body Temperature: C. R. Bell*: London School of Hygiene and Tropical Medicine, University of London, Keppel Street, (Gower Street), London, W.C.1., England.

Discriminability and Scaling of Linear Extent: C. Douglas Creelman*: Subcommittee on Noise Research Center, 327 South Alvarado Street, Los Angeles, California 90057.

Visual Field and the Letter Span: Herbert F. Crovitz* and H. Richard Schiffman: VA Hospital, Fulton Street and Erwin Road, Durham, North Carolina 27705.

Expected Value and Response Uncertainty in Multiple-Choice Decision Behavior: David M. Messick* and Amnon Rapoport: Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina.

Task Predictability in the Organization, Acquisition, and Retention of Tracking Skill: Don Trumbo,* Merrill Noble, Kenneth Cross, and Lynn Ulrich: Department of Psychology, Anderson Hall, Kansas State University, Manhattan, Kansas 66504.

Isolation Effects When Paired Associates are Presented Serially: Slater E. Newman* and G. Alfred Forsyth: Department of Psychology, North Carolina State College, Raleigh, North Carolina.

Interlist Response Meaningfulness and Transfer Effects under the A-B, A-C Paradigm: L. R. Goulet*: Department of Psychology, Saint Louis University, 221 North Grand Boulevard, Saint Louis 3, Missouri.

Reminiscence and Forgetting in a Runway: Winfred F. Hill,* Albert Erlebacher, and Norman E. Spear: Department of Psychology, Northwestern University, Evanston, Illinois 60201.

On the Different Effects of Random Reinforcement and Presolution Reversal on Human Concept Identification. Solon B. Holstein and David Premack*: Department of Psychology, University of Missouri, Columbia, Missouri.

Optimal Responding in Multiple-Cue Probability Learning: Cameron R. Peterson,* Kenneth R. Hammond, and David A. Summers: Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan, P.O. Box 618, Ann Arbor, Michigan.

* Asterisk indicates author for whom the address is supplied.

Effects of List Length in the Ebbinghaus Derived-List Paradigm: Robert K. Young, David T. Hakes,* and R. Yale Hicks: Department of Psychology, Mezes Hall 211, University of Texas, Austin 12, Texas.

Short-Term Perceptual Recognition Memory for Tachistoscopically Presented Nonsense Forms: Richard A. Steffy and Charles W. Eriksen*: Department of Psychology, University of Illinois, Urbana, Illinois 61803.

Size Cues and the Adjacency Principle: Walter C. Gogel*: Spatial Perception Section, Civil Aeromedical Research Institute, P.O. Box 1082, Oklahoma City, Oklahoma 73101.

Unlearning under Conditions of Successive Interpolation: Leo Postman*: Department of Psychology, University of California, Berkeley, California 94720.

Effect of Pairing a Stimulus with Presentations of the UCS on the Extinction of an Avoidance Response in Humans: Robert K. Banks*: Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

Unavailability and Associative Loss in RI and PI: John Ceraso* and Ann Henderson: Department of Psychology, Graduate School of Education, Yeshiva University, 110 West 57th Street, New York, New York 10019.

Acquisition-Transfer, and Retention of S-R_1, S-R_2 Associations as a Function of S-R_1, S-R_2 Pattern: Richard Popp and James F. Voss*: Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

Fixation Time as a Function of Stimulus Uncertainty: James Allison*: Department of Psychology, Indiana University, Bloomington, Indiana.

Information and Reaction Time for "Naming" Responses: Robert E. Morin,* Andrew Konick, Nola Troxell, and Sandra McPherson: Department of Psychology, Kent State University, Kent, Ohio.

Extended Training Extinction Effect under Massed and Spaced Extinction Trials: David Birch*: Department of Psychology, University of Michigan, Ann Arbor, Michigan.

Selection Strategies in Concept Attainment as a Function of Number of Persons and Stimulus Display: Patrick R. Laughlin*: Department of Psychology, Northwestern University, Evanston, Illinois 60201.

Running Speed in the Rat as a Function of Shock Level and Competing Responses: George A. Cicala* and J. R. Corey: Department of Psychology, University of Delaware, Newark, Delaware.

Effect of Delayed Punishment of an Immediately Rewarded Response in Humans: R. K. Banks* and M. Vogel-Sprott: Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

Further Study of Effects of Low Drive States on Competing Responses: Charles Y. Nakamura* and William E. Broen, Jr.: Department of Psychology, University of California, Los Angeles, California 90024.

Generalization to Similar and Opposite Words: Richard K. Olson*: Department of Psychology, University of Oregon, Eugene, Oregon 97403.

Effects of Consummatory Response Punishment in Spatial Discrimination Learning and Response Fixation: Charles H. Koski and Leonard E. Ross*: Department of Psychology, West Johnson and Charter Streets, Madison, Wisconsin 53706.

Sequential Dependencies in Children's Probability Learning: Richard S. Bogartz*: State University of Iowa, Institute of Child Behavior and Development, Iowa City, Iowa 52240.

Effects of Delay and Mode of Presentation of Extra Cues on Pursuit-Rotor Performance: Lawrence Karlin*: Department of Psychology, New York University, Washington Square, New York, New York 10003.

Approximations to English (AE) and Short-Term Memory: Construction or Storage?: Roy Lachman* and Abigail V. Tuttle: Department of Psychology, State University of New York, Buffalo, New York 14214.

A Reinterpretation of One Form of Backward and Forward Masking in Visual Perception: Charles W. Eriksen* and James F. Collins: Department of Psychology, University of Illinois, Urbana, Illinois 61803.

The Use of Negative Slope Transformations of Knowledge of Results on a Simple Motor Response: Donald A. Schumsky*: Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221.

A Comparison of Paired-Associate Transfer Effects between the A-B, C-A and A-B, B-C Paradigms: L. R. Goulet* and A. Barclay: Department of Psychology, Saint Louis University, 221 North Grand Boulevard, Saint Louis 3, Missouri.

Spatial and Temporal Determinants of Visual Backward Masking: Robert W. Sekuler*: Department of Psychology, Brown University, Providence, Rhode Island 02912.

Interaction of Time Uncertainty and Relative Signal Frequency in Determining Choice Reaction Time: Paul Bertelson* and Jacques Barzeele: Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

Sonic and Ultrasonic Equal Loudness Contours: John F. Corso* and Murray Levine: Department of Psychology, State University of New York, Cortland, New York.

Size and Structure of Grammatical Units in Paired-Associate Learning at Two Age Levels: James G. Martin* and Robert L. Jones: Division of Education and Psychology, Chico State College, Chico, California.

Discrimination Learning as a Function of Varying Pairs of Sucrose Rewards: Roger W Black*: Department of Psychology, State University of Iowa, Iowa City, Iowa.

A Reversed Partial-Reinforcement Effect: William B. Pavlik* and Peter L. Carlton: Department of Psychology, Rutgers, New Brunswick, New Jersey 08903.

Sex Differences in Adaptation of the GSR under Repeated Applications of a Visual Stimulus: H. D. Kimmel* and Ellen Kimmel: Department of Psychology, University of Florida, Gainesville, Florida.

Detection of a Visual Signal With Low Background Noise: An Experimental Comparison of Two Theories: Raymond H. Hohle*: Institute of Child Behavior and Development, State University of Iowa, Iowa City, Iowa 52240.

Information and Certainty: Kenneth B. Little* and Larry M. Lintz: Department of Psychology, University of Denver, University Park, Denver, Colorado 80210.

Differential Adaptation of Anxious and Nonanxious Children in Instrumental Escape Conditioning: Ronald K. Penney*: Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

Proactive Interference as a Function of Time between Tests: Lloyd R. Peterson* and Antoinette Gentile: Department of Psychology, University of California, Berkeley 4, California.

Effects of Distribution of Practice of the Acquisition Pairs upon Mediating Responses and Response Availability: Margaret Jean Peterson* and Sarah M. Jamison, Institute of Human Learning, University of California, Berkeley, California 94720.

Effect of Size and Location of Informational Transforms upon Short-Term Retention: Michael I. Posner* and Ellen Rossman: Department of Psychology, University of Wisconsin, West Johnson and Charter Streets, Madison, Wisconsin 53706.

Signal-Detection Theory and Short-Term Memory: Bennet B. Murdock, Jr.*: Department of Psychology, 13 McAlester Hall, University of Missouri, Columbia, Missouri 65202.

Adjustment to New Reward: Simultaneous and Successive Contrast Effects: Norman E. Spear* and Winfred F. Hill: Department of Psychology, Rutgers, New Brunswick, New Jersey 08903.

Intrasession Decrements in the Performance of the Classically Conditioned Eyelid Reflex: Willard N. Runquist* and William R. Muir: Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

Grammatical Agreement and Set in Learning at Two Age Levels: James G. Martin,* Judy R. Davidson, and Myrna L. Williams: Division of Education and Psychology, Chico State College, Chico, California 95926.

Effects of Number and Percentage of Rewarded Trials on the Acquisition and Extinction of Lever Pressing Using a Discrete-Trial Procedure: John J. Porter* and James J. Hug: Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53211.

Intrasession Adaptation and Intersession Extinction of the Components of the Orienting Response: Charles R. Galbrecht,* Roscoe A. Dykman, William G. Reese, and Tetsuko Suzuki: Medical Center, Department of Psychiatry, University of Arkansas, Little Rock, Arkansas.

A Psychophysical Analysis of the Odor Intensity of Homologous Alcohols: Trygg Engen*: Walter S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912.

Effects of the Visual Field upon Perception of Change in Spatial Orientation: Norman L. Corah*: Department of Psychology, Washington University, St. Louis 30, Missouri.

Short-Term Recognition Memory for Single Digits and Pairs of Digits: Donald A. Norman,* and Wayne A. Wickelgren: Center for Cognitive Studies, Harvard University, 61 Kirkland Street, Cambridge 38, Massachusetts.

Anagram Solution Times: A Function of Individual Differences in Stored Digram Frequencies: M. E. Tresselt and M. S. Mayzner*: Department of Industrial Engineering and Operations Research, New York University, University Heights, Bronx, New York 10053.

Cutaneous Sound Localization: George A. Gescheider*: Department of Psychology, Hamilton College, Clinton, New York 13323.

Improvement in Recall on Unreinforced Recall Trials: Rose Greenbloom and Gregory A. Kimble*: Department of Psychology, Duke University, Durham, North Carolina 27706.

Effects of Interval between Successive Numbers and Pattern in Visual Learning: Charles P. Thompson*: Department of Psychology and Philosophy, University of Wyoming, Laramie, Wyoming 82071.

Speed of Nonreinforced Running Response Following Increasing and Decreasing Orders of Sucrose Concentrations: Melvin H. Marx* and David C. Edwards: Department of Psychology, University of Missouri, Columbia, Missouri 65202.

Effect of Contextual Associations upon Selective Reaction Time in a Numeral-Naming Task: Bert Forrin*: Department of Psychology, Kent State University, Kent, Ohio 44240.

Prediction of Responses to Multidimensional from Responses to Unidimensional Stimuli: D. W. J. Corcoran*: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

Changes in Perceived Size of Angle as a Function of Orientation in the Frontal Plane: Paul Weene and Richard Held*: Department of Psychology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

A Replication of Reaction Time to Stimuli Masked by Metacontrast: Keith Harrison* and Robert Fox: Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

Anagram Solution Times: A Function of Multiple Solution Anagrams: M. S. Mayzner* and M. E. Tresselt: Department of Industrial Engineering and Operations Research, New York University, University Heights 53, New York.

Doublet Effects in the Verbal Maze during Acquisition and Relearning: Charles P. Thompson*: Department of Psychology and Philosophy, University of Wyoming, Laramie, Wyoming 82071.

Effect of Size on Visual Slant: Robert B. Freeman, Jr.*: Department of Psychology, Pennsylvania State University, Burrowes Building, University Park, Pennsylvania 16802.

Reminiscence in Short-Term Retention: Lloyd R. Peterson*: Department of Psychology, University of California, Los Angeles, California 90024.

Response Transfer as a Function of Verbal-Association Strength: Lynn K. Brown,* James J. Jenkins, and Joyce Lavik: Student Counseling Bureau, Eddy Hall-101, University of Minnesota, Minneapolis 14, Minnesota.

Abstraction in Verbal Paired-Associate Learning: Barbara S. Musgrave* and Jean Carl Cohen: Department of Psychology, Smith College, Northampton, Massachusetts.

Visual and Proprioceptive Adaptation to Optical Displacement of the Visual Stimulus: John C. Hay* and Herbert L. Pick, Jr.: Psychology Department, Smith College, Northampton, Massachusetts.

Cognitive Factors in Impairment: A Neuropsychological Study of Divided Set: Robert B. Malmo*: Allan Memorial Institute, McGill University, 1025 Pine Avenue, W., Montreal 2, Quebec, Canada.

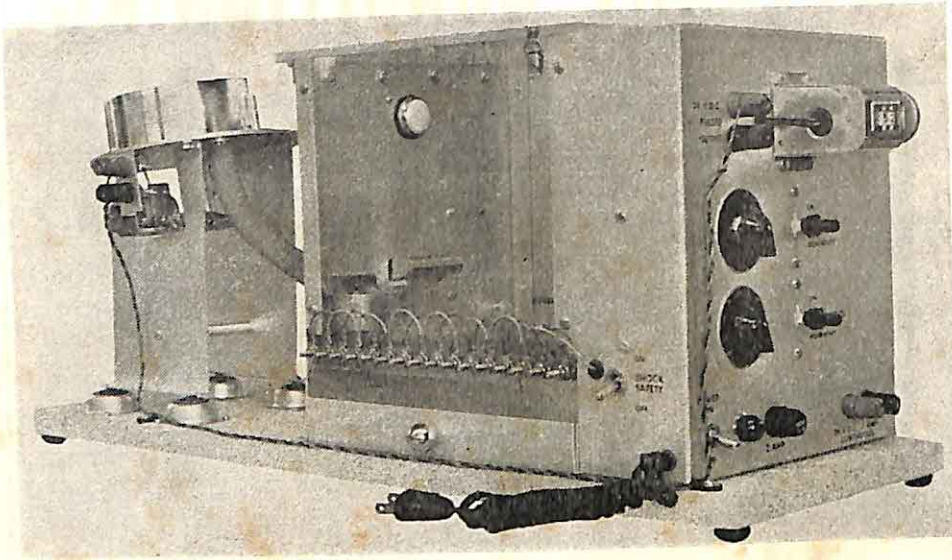
Motivational Factors in Short-Term Retention: Bernard Weiner* and Edward L. Walker: Department of Psychology, University of Minnesota, Minneapolis 55, Minnesota.

Latency and Magnitude of GSR as a Function of Interstimulus Interval: Arno F. Wittig* and Delos D. Wickens: Department of Psychology, Ball State Teachers College, Munsie, Indiana 47306.

Anxiety and Eyelid Conditioning: Mark Ominsky and Gregory A. Kimble*: Department of Psychology, Duke University, Durham, North Carolina 27706.

Perceived Duration as a Function of Auditory Stimulus Frequency: Austin Jones* and Marilyn MacLean: Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania.

MODEL A-105



Student Lab or Demonstration Box Animal Cage with Many Features by Scientific Prototype

This new Scientific Prototype animal cage combines cage, pellet feeder, power supply, light and shock source intensity controls, and counter. The unit is completely wired, mounted on a base, and ready to operate.

The rat box is approximately 9 × 8⅝ × 8. Grids are stainless steel. Top and sides are ¼″ clear Plexiglass. The stainless steel pedal lever is counter-balanced to permit sensitivity adjustments. Dropping tray is aluminum anodized.

Of special interest is the feature of self-contained cue light and shock source *intensity controls.*

Price . . . $259.00 F.O.B. N.Y.C. . . . quantity discounts. Full information available.

Behavior Analysis Apparatus . . . operant conditioning kits, tachistoscopes, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital, modular test equipment; and special systems

SCIENTIFIC PROTOTYPE MFG., CORP.
623 W. 129th STREET • NEW YORK, N. Y.
TELEPHONES: (212) MO 6-4674 and 4789

Journal of Experimental Psychology

Vol. 69, No. 3 March 1965

MOMENTS OF AREA AND OF THE PERIMETER OF VISUAL FORM AS PREDICTORS OF DISCRIMINATION PERFORMANCE [1]

LEONARD ZUSNE [2]

Purdue University

2 experiments were performed to evaluate the usefulness of the 2nd, 3rd, and 4th moments of distribution of the area and of the perimeter of random shapes as predictors of performance in a discrimination task. Repeated measures of RT were taken on 44 Ss who responded to 80 6-choice oddity-type problems defined by the presence or absence of a difference in the 3 moments. Both moments of area and of the perimeter proved to be significant ($p < .01$) predictors of performance. Computer method for the computation of moments is presented, and the difference between the 2 types of moments, methodological problems, their solution, and the advantages of moments as compared to other form measures are discussed.

Between 70 and 80 physical measures of visual form have been defined and used in form-perception experiments since 1948 when the formulation of information theory gave impetus to the development of metrics of visual form. With a few notable exceptions, this profusion of measures has not led to a definition of the limits of the multidimensionality of form, much less to the establishment of a few basic parameters.

Four of these measures—information content, linearity of the contour, area, and rotation—belong to one class. They differ from the rest in that they do not directly describe the manner in which form contours are distributed in space. Such a description has been the main difficulty so far, and it accounts for the large number of measures that make up the remaining group. Among these, the measures of compactness, symmetry, and elongation have consistently emerged as important predictors of performance in a variety of perceptual tasks, experimental conditions, and with both human and animal *S*s (e.g., Adams, Fitts, Rappaport, & Weinstein, 1954; Arnoult,

[1] This investigation was supported in part by Research Grant M-4532 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, and in part by the Purdue Research Foundation, Purdue University. The author is thankful to William H. Payne for having written the computer program and to Kenneth M. Michels, Project Director, for his help, constructive criticism, and encouragement during the course of this investigation.

[2] Now at the University of Tulsa.

1960; Attneave, 1957; Bitterman, Krauskopf, & Hochberg, 1954; Casperson, 1950; Fitts, Weinstein, Rappaport, Anderson, & Leonard, 1956; Lamar, Hecht, Shlaer, & Hendley, 1947; Zusne & Michels, 1962a, 1962b). It appears now that the definitions of these three parameters can be improved, indeed, that these and other related measures may all be brought to a common denominator.

In statistics, a unidimensional sampling distribution is specified by stating its mean, variance, skewness, kurtosis, and other higher moments. The area of a shape may be thought of as a two-dimensional distribution of its elements, i.e., small units of area. Such a distribution may be specified by stating its moments of area. When it is realized that the three physical form measures just mentioned are related to the second (compactness = variance), third (skewness = symmetry), and fourth (kurtosis = elongation) moments of a sampling distribution, the investigation of moments of area acquires additional interest.

Alt (1962) appears to be the only investigator to have used moments of area to specify two-dimensional distributions, but his work does not deal with the behavioral correlates of moments. The overlap measures developed by Monty and Boynton (1962), while not explicitly recognized as such by the authors, are related to moments of area.

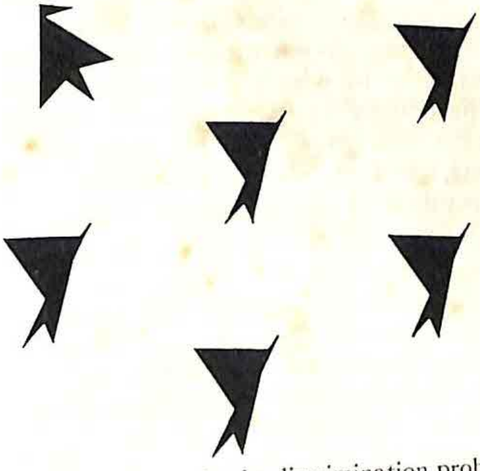


FIG. 1. Example of a discrimination problem used in Exp. I and II. (The *Ss* saw bright figures on a dark background.)

A pilot study showed that when 50 *Ss* were given the task of drawing a straight line through each of 100 eight-sided random shapes so as to maximize the symmetry of the two halves of the shape, the location of the line coincided or was very close to that axis about which the moments of area were either at their minimum or maximum values. The two experiments described below were designed to further test the utility of moments as predictors of performance.

EXPERIMENT I

Method

Method of computing moments of area.—The computation of moments of area amounts to double integration over the rectangle containing the shape (two sides of which are the coordinate axes). Using the Lebesgue integral, for one variable X moments are defined as

$$U_k = \iint_R (X - \bar{X})^k dA,$$

where k is the order of the moment, R is the region within the boundary of the form, X is the distance of an incremental area dA from the origin, and $\bar{X}$ is the distance of the centroid from the origin.

Since the computation of moments is a very time-consuming task, a computer program was written for this purpose. In general, the basic unit of which the various moments of area are composed is the distance-area product, i.e., a unit area multiplied by its distance from a coordinate axis. Where the shape is positioned in the coordinate system is immaterial for the purposes of computer operations. A shape for which moments of area are to be computed is divided into small, discrete area units. These units are coded by 1 punches (nonarea by 0 punches) in IBM cards. The computer operation amounts to summing these unit areas, multiplying them by their respective distances from the coordinate axes, and evaluating the computational formulae for the moments. Looking at it in another way, the shape is projected unto

an axis (unidimensional collapse) and moments of distribution are computed on the resulting frequency distribution polygon of the unit areas of the shape. Rotating the shape changes the form of this distribution polygon, hence also the values of the computed moments. The computer program was written for the first four moments (later augmented to first eight) to be computed at each 5° of rotation of the shape.

Subjects.—The *S*s were 11 male and 11 female students enrolled in an introductory psychology course at Purdue University.

Stimuli and apparatus.—A pool of 100 eight-sided random shapes was constructed according to Method I of Attneave and Arnoult (1956). These were the 100 shapes used in the previously mentioned pilot study. The shapes were equated on area (19.22 sq. in.), divided into units of .1-sq. in. area, and the first four moments were computed for each shape. The stimuli were selected from this pool of shapes, as described below. The stimulus shapes were cut from black paper and arranged in problems as shown in Fig. 1; i.e., two shapes, one of which was replicated five times. The problems were photographed on 35-mm. negative film and each film frame was mounted as a 2-in. projector slide.

A Kodak Carousel slide projector projected the slides on a 15-in. square Albanene paper screen mounted in the front wall of a viewing booth. The *S* sat facing the other side of the screen, which was about 2½ ft. away. The booth was equipped with a response board having six button switches arranged in the same way as the six shapes in a problem. Reaction time was recorded by a Standard Electric timer in one-hundredths of a second. Located between *S*'s booth and the projector was *E*'s control board equipped with the necessary controls for the presentation of the stimuli and for closing the appropriate circuits.

Experimental design.—The high correlation of .91 between the second and the fourth moments, computed about the perceptual axis of symmetry defined in the above-mentioned pilot study, made it impossible to vary these two parameters independently. However, the function of the second moment (i.e., the plot of its values computed in successive rotations through a complete cycle) is symmetrical about its maximum or minimum point; in addition, the graph of the *x* function of the second moment is a mirror image of the graph of the *y* function. For this reason the two second moments sum to a constant in any rotation of the shape. This rotation-independent measure of the second moment did not correlate too highly with either the third or the fourth moments, which determined its use in the experiment.

Statistically, the design was a $2 \times 2 \times 2 \times 5$ factorial design with repeated measures taken on all factors. The first three factors were: sum of the second moments (second moment constant), the third moment about the perceptual axis of symmetry, and the fourth moment about the same axis. (To make moments more concrete, they will be provisionally referred to in the following as areal dispersion or Da, areal symmetry or Sa, and areal kurtosis or Ka.) Each of these factors occurred as either "present" or "absent" in terms of a difference in their values between the two shapes that constituted a discrimination problem. The fourth factor, nested under the first three, was problems within the eight experimental conditions. All factors were considered to be fixed.

Forty six-choice, oddity-type discrimination problems (5 for each experimental condition) plus 6 practice problems were constructed as follows: From the 100 shapes, pairs of shapes were selected to represent the eight possible combinations of presence or absence of a difference in the value of the three moments. The size of the difference was about one-fourth of the total range of the respective moment for the 100 shapes.

The shapes always appeared with their perceptual axes of symmetry in the vertical orientation. The location of the "different" shape was randomized. Each problem was presented once to each *S*, the order of presentation also being random, except that the six practice problems always preceded the series. This order was reversed for one-half of the *S*s of each sex.

The *S*s were tested individually. The instructions directed *S* to find the "different" shape among the six and to press the button on the response board which corresponded to its location. The *S* was informed that his RT would be measured and that he was to respond as fast as possible without making an error. The *S* was also instructed to use only his writing hand for responding. The *E* recorded the RT and proceeded with the next slide.

Results

Unweighted means analysis on sex differences showed that this factor did not affect the RT significantly. Except for the practice problems, there was no learning over blocks of problems in any of the *S*s. The practice problems were not used in the computations.

TABLE 1

Analysis of Variance of Log (1 + RT) to Discrimination Problems in Exp. I and II

Source	df	Exp. I			Exp. II		
		MS	Error	*F*	*MS*	Error	*F*
Between *S*s	21						
Within *S*s	858						
Second moment constant (A)	1	.04014	.00354	11.34**	.07870	.00554	14.21**
Third moment (B)	1	.03536	.00446	7.93*	.03022	.00329	9.18**
Fourth moment (C)	1	.00342	.00572	.60	.07990	.00556	14.37**
Probl. within $A_1B_1C_1$[a]	4	.00104	.00492	.21	.00622	.00405	1.53
Probl. within $A_1B_1C_0$	4	.02315	.00232	9.98**	.03756	.00363	10.34**
Probl. within $A_1B_0C_1$	4	.00511	.00254	2.01	.03237	.00496	6.59**
Probl. within $A_1B_0C_0$	4	.02621	.00452	5.79**	.00745	.00302	2.46
Probl. within $A_0B_1C_1$	4	.10129	.00367	27.59**	.00882	.00383	2.30
Probl. within $A_0B_1C_0$	4	.02422	.00249	9.72**	.05894	.00586	10.06**
Probl. within $A_0B_0C_1$	4	.01332	.00409	3.25*	.00619	.00222	2.79*
Probl. within $A_0B_0C_0$	4	.22219	.00420	52.90**	.03540	.00454	7.79**
A × B	1	.01161	.00311	3.73	.02009	.00217	9.26**
A × C	1	.00236	.00247	.96	.04553	.00588	7.74**
B × C	1	.33097	.00719	46.03**	.00365	.00411	.89
A × B × C	1	.00594	.00312	1.90	.08536	.00336	25.40**

[a] 1 = difference present, 0 = difference absent.
* $p < .05$.
** $p < .01$.

A plot of the distribution of RTs showed positive skewness. A log (1 + RT) transformation was therefore applied to the data. A summary analysis of variance of the transformed RTs appears in Table 1. Two main conclusions can be drawn from this analysis. First, the presence or absence of a difference in either Da or Sa significantly affected *S*s' RT ($p < .01$). No such effect appeared in the case of Ka. In addition, the higher was the order of the moment the smaller a proportion of the total variation was accounted for by it. Second, in six out of the eight experimental conditions problems significantly affected the results, indicating that heterogeneous samples of problems had been constructed.

Of the four interactions only one was statistically significant. Analysis of simple main effects showed that the interaction between Sa and Ka occurred as follows. Performance was best when a difference in only one of the moments was present. In this situation Sa was a significantly ($p < .01$) better predictor than Ka. When differences in both moments

TABLE 2

Mean Reaction Time (in Sec.) to Discrimination Problems in Eight Experimental Conditions of Exp. I

	Experimental Cond.[a]							
	$A_0B_0C_0$	$A_0B_1C_1$	$A_1B_0C_0$	$A_1B_1C_1$	$A_0B_0C_1$	$A_1B_1C_0$	$A_1B_0C_1$	$A_0B_1C_0$
Mean RT	1.61	1.45	1.43	1.35	1.29	1.20	1.19	1.19

Note.—Underlined conditions do not differ significantly (conservative Newman-Keuls test).
[a] A = second moment constant, B = third moment, C = fourth moment, 1 = difference present, 0 = difference absent.

were present, discrimination deteriorated considerably ($p < .01$), indicating interference. Performance was poorest, however, when differences in both moments were absent ($p < .01$).

Table 2 shows the mean RTs for the eight experimental conditions. It is interesting to note that a difference in Da or Ka alone resulted in a considerably poorer performance than when differences in both moments were present simultaneously, while a difference in Sa alone led to best performance and the addition of either Da or Ka or both deteriorated it.

Discussion

The probable reason why the main effect of Ka did not turn out to be statistically significant is discussed under Exp. II. The interaction between Sa and Ka was highly significant, however. The pattern of interaction indicated interference between these two variables. Whether the observed interference was mutual or unilateral is not clear, and further experiments will have to be conducted to establish this point as well as the exact reason for the interference.

It was noted that in each of the six experimental conditions showing heterogeneity among problems it was due to only one or at most two markedly deviant problems. It was established that of the various possible sources of heterogeneity only two were actually involved. One source was the algebraic sign of Sa. While the absolute values of Sa of a shape in 0 and 180° rotations are the same, they differ in algebraic sign. Thus in some problems with no difference in Sa the differing algebraic signs introduced a discriminable clue in the form of discriminable directionality of the two shapes.

The characteristic of the second group of deviant problems can be illustrated by comparing, for instance, a square with the same square from which a short, pointed spike protrudes. The two squares are clearly discriminable, yet, because moments of area are insensitive to small, protruding areas, the computed moments of the two squares hardly differ.

The first source of heterogeneity of problems can be eliminated by simply rotating one of the two shapes 180°. Two possible ways of eliminating the "spike" problem suggested themselves, namely the use of moments of the perimeter instead of moments of area and/or the use of more than the first four moments to specify problems. These possibilities were explored in a second experiment.

Experiment II

Although moments of area describe the distribution of the area enclosed by the contour of a shape, this area is really "empty" in that it contains no information—all information in a simple, two-dimensional, black-and-white polygon is contained in its contour. This has been shown by Attneave and Arnoult (1956) and is indicated in eye-movement studies (Brandt, 1945; Buswell, 1935). A study of eye movements in response to random and geometric polygons (Zusne & Michels, 1964) has shown that in such stimuli it is the outlines of the shape that are scanned, especially the intricate, interesting, or important portions of it, while the enclosed area is almost completely ignored. It seemed indicated, therefore, to test the utility of moments of the perimeter, i.e., moments of shape weighted entirely along its contour, leaving the enclosed area out of consideration.

The above-mentioned "spike" phenomenon also suggested the possibility that heterogeneous problems had been constructed because not all of the relevant moments of shape had been specified. This brings up what is known in mathematics as the "problem of moments," which is simply this: if all moments of a distribution are given, is this distribution thereby uniquely specified? The answer to this problem has not yet been given, and it is well to note that

TABLE 3

Intercorrelations among Moments of Area and Moments of the Perimeter of 100 Random Shapes

	$U2a_c$	U2a	U3a	U4a	$U2p_c$	U2p	U3p	U4p	U5p	U6p	U7p
$U2a_c$	—	.53	.41	.62	.88	.36	.08	.32	.14	.30	.27
U2a		—	.56	.91	.31	.66	.14	.62	.22	.55	.32
U3a			—	.46	.34	.76	.23	.79	.33	.78	.51
U4a				—	.42	.44	.11	.46	.20	.45	.28
$U2p_c$					—	.39	.07	.34	.14	.32	.23
U2p						—	.23	.94	.34	.86	.48
U3p							—	.27	.94	.29	.74
U4p								—	.40	.97	.56
U5p									—	.44	.87
U6p										—	.61
U7p											—

Note.—Small letters designate a moment of area or of the perimeter, number indicates the order of the moment, and c is the second moment constant.

it has been raised for one-dimensional distributions only. In a concrete case the specification of "all moments" is effectively curtailed by the capacity of even the largest computers, which limits the investigator to the specification of "more" rather than "all" moments.

Method

Subjects.—Twenty-two male students enrolled in an introductory psychology course at Purdue University served as *S*s.

Stimuli and apparatus.—The same pool of shapes was used as in Exp. I. In punching a shape only its contour was represented by 1 punches, while area both outside the shape and inside it was coded by 0 punches. An amplified but essentially the same computer program was used to compute the first eight moments of the perimeter for the 100 shapes.

The method of stimulus presentation and the apparatus used were the same as in Exp. I.

Experimental design.—Table 3 shows that all even moments of the perimeter were highly correlated, as were all odd moments. This made the construction of problems using differences in the size of all moments both impossible and unnecessary. As before, the second moment constant did not correlate too highly with any of the other moments. This measure plus the third and fourth moments of the perimeter computed about the perceptual axis of symmetry were used to define the experimental conditions. In the following, these measures will be referred to as perimetric dispersion or Dp, perimetric symmetry or Sp, and perimetric kurtosis or Kp, respectively.

Forty discrimination problems were constructed according to the plan used in Exp. I. The statistical design, instructions, and experimental procedure were identical to those used in Exp. I.

Results

An analysis of variance was performed on log (1 + RT). A summary of the ANOVA appears in Table 1.

TABLE 4

Mean Reaction Time (in Sec.) to Discrimination Problems in Eight Experimental Conditions of Exp. II

	Experimental Cond.[a]							
	$A_0B_1C_0$	$A_0B_0C_0$	$A_0B_1C_1$	$A_1B_1C_0$	$A_1B_0C_1$	$A_1B_1C_1$	$A_1B_0C_0$	$A_0B_0C_1$
Mean RT	1.53	1.51	1.43	1.40	1.39	1.26	1.25	1.22

Note.—Underlined conditions do not differ significantly (conservative Newman-Keuls test).
[a] A = second moment constant, B = third moment, C = fourth moment, 1 = difference present, 0 = difference absent.

In contrast to Exp. I, the main effects of all three moments were found to be statistically significant ($p < .01$), with no decrease in the size of the F ratio with increased order of the moment. Similar to Exp. I, problems within experimental conditions contributed significantly to the total variation; the overall heterogeneity of problems, however, was less pronounced than in Exp. I.

Of the four interactions only one was not statistically significant. Analysis of simple main effects showed the following patterns. As long as a difference in Dp was present, the presence or absence of a difference in Sp did not have any appreciable effect. When no difference in Dp was present, the presence of a difference in Sp significantly increased RT ($p < .01$).

The interaction of the two even moments, however, was quite different. The RTs did not differ significantly among themselves when a difference in either Dp, or Kp, or both was present. Performance did deteriorate significantly ($p < .01$) when a difference in both moments was absent. Table 4 shows the mean RTs for the eight experimental conditions. When this table is compared with Table 2, a very definite, systematic relationship is seen: both the easiest and most difficult conditions in Exp. I were the most difficult in Exp. II, while conditions of medium difficulty were the same in both experiments.

Discussion

The systematic shift in the degree of difficulty between the experimental conditions of the two experiments may be explained, partially at least, by the difference between the configurational properties represented by moments of area and moments of the perimeter. Examples of shapes with the highest and the lowest values of moments of area and of the perimeter in the pool of 100 shapes are shown in Fig. 2. While Sa is a direct

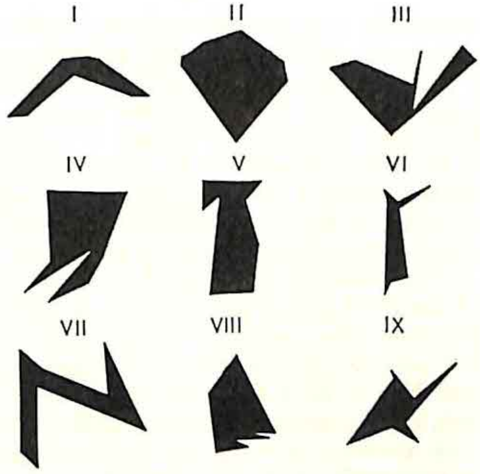


FIG. 2. Examples of shapes high and low on the three measures of moments used in the two experiments. (These shapes had the highest and the lowest values of these moments among the 100 shapes in the original sample. I—high Da and Ka; II—low Da and Dp; III—high Sa; IV—low Sa; V—low Ka and Kp; VI—high Dp; VII—high Sp; VIII—low Sp; IX—high Ep.)

index of the symmetry of the two halves of a shape about its perceptual axis of symmetry, Sp expresses the balance between the two halves in terms of the length of the perimeter in each half. When the perimeter is compressed in a small area, it produces perimetric asymmetry without necessarily producing areal asymmetry. This lack of relationship is reflected in the rather low correlation coefficient between these two measures ($r = .23$). In the experimental condition where a difference in only Sp was present, the asymmetrical distribution of the perimeter in the "different" shape was not easily noticeable, which probably accounts for the fact that this condition turned out to be the most difficult one.

The dispersion of the perimeter in terms of Dp, however, is closely related to the dispersion of the area of a shape ($r = .88$), and once again proved to be a valuable predictor of performance. The Kp measure, on the other hand, is not too highly related to Ka ($r = .46$). Shapes that are leptokurtic in terms of Kp tend to have one or more contour lines in a vertical position. Such a shape

turned 90° is platikurtic. Thus a difference in Kp may show itself as a difference in the directionality of the shape, which is more readily perceived than an equal difference between two shapes in terms of Ka. The elongation described by the latter, being a sort of average of all portions of the shape, apparently needs much larger differences to be perceived.

The difference between the two types of the third and fourth moments is quite probably also related to the fact that the one significant interaction (between Sa and Ka) in Exp. I turned out to be the only nonsignificant interaction in Exp. II. The present experiments, however, do not provide sufficient grounds to hypothesize about the exact nature of the relationship between the third and the fourth moments.

The "spike phenomenon" could not be completely eliminated by the use of moments of the perimeter, but its magnitude was reduced. It appears, therefore, that in future experiments differential weighting of the perimeter would have to be resorted to by weighting perceptually conspicuous portions of the perimeter more than its other parts.

The investigation reported in this paper was only an initial step in the evaluation of moments as form parameters, and many questions remain to be answered and many problems solved. Nevertheless, some of the advantages of moments as compared with the form measures of this category used so far are apparent even now: (*a*) Instead of a large number of unrelated measures a small number of measures having the same mathematical basis is used. (*b*) The possible uniqueness of representation has already been mentioned. Such a representation, if a fact, means that moments exhaust all possible configurational aspects of two-dimensional visual displays. (*c*) Moments permit the quantification of form on continua rather than in a discrete or dichotomous fashion. (*d*) Plotting the functions of moments through a complete cycle of rotation permits fast, easy, and quantified comparison of shapes on the various characteristics represented by the moments. (*e*) Finally, moments have the advantage of possessing a relatively broad interdisciplinary basis: the concept of moments is a well-established one in mathematics, statistics, physics, and the several branches of engineering.

REFERENCES

ADAMS, O. S., FITTS, P. M., RAPPAPORT, M., & WEINSTEIN, M. Relations among some measures of pattern discriminability. *J. exp. Psychol.*, 1954, **48,** 81–88.

ALT, F. L. Digital pattern recognition by moments. In G. L. Fischer, Jr., et al. (Eds.), *Optical character recognition.* Washington, D.C.: Spartan Books, 1962.

ARNOULT, M. D. Prediction of perceptual responses from structural characteristics of the stimulus. *Percept. mot. Skills*, 1960, **11,** 261–268.

ATTNEAVE, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53,** 221–227.

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53,** 452–471.

BITTERMAN, M. E., KRAUSKOPF, J., & HOCHBERG, J. E. Threshold for visual form: A diffusion model. *Amer. J. Psychol.*, 1954, **67,** 205–219.

BRANDT, H. F. *The psychology of seeing.* New York: Philosophical Library, 1945.

BUSWELL, G. T. *How people look at pictures.* Chicago: Chicago Univer. Press, 1935.

CASPERSON, R. C. The visual discrimination of geometric form. *J. exp. Psychol.*, 1950, **40,** 668–686.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N. S., & LEONARD, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1956, **51,** 1–11.

LAMAR, E. S., HECHT, S., SHLAER, S., & HENDLEY, C. D. Size, shape, and contrast in detection of targets by daylight vision: I. Data and analytical comparison. *J. Opt. Soc. Amer.*, 1947, **37,** 531–545.

MONTY, R. A., & BOYNTON, R. M. Stimulus overlap and form similarity under suprathreshold conditions. *Percept. mot. Skills*, 1962, **14,** 487–498.

ZUSNE, L., & MICHELS, K. M. Geometricity of visual form. *Percept. mot. Skills*, 1962, **14,** 147–154. (a)

ZUSNE, L., & MICHELS, K. M. More on the geometricity of visual form. *Percept. mot. Skills*, 1962, **15,** 55–58. (b)

ZUSNE, L., & MICHELS, K. M. Nonrepresentational shapes and eye movements *Percept. mot. Skills*, 1964, **18,** 11–20.

(Received January 27, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 221–223

STIMULUS CORRELATES OF VISUAL PATTERN DISCRIMINATION BY HUMANS: AREA AND CONTOUR [1]

V. J. POLIDORA

Wisconsin Regional Primate Research Center

This experiment was a replication, with human Ss, of an earlier one with rhesus monkeys. Human proficiency of discriminating visual metric patterns was virtually perfect and less directly related to the area and contour dimensions of these patterns than was the performance of monkeys on the same discrimination problems.

This research program is concerned with identifying the physical stimulus dimensions utilized by primates in discriminating visual patterns. The initial report (Polidora & Thompson, 1964) established that the disparity between (*a*) the areas and (*b*) the contours of the two patterns which comprised a simultaneous discrimination problem were both direct, monotonic determinants of discriminative performance of monkeys. Further, the contributions of the area and contour dimensions were shown to be independent and additive. To determine the extent to which these same dimensions relate also to human discriminative performance, the present experiment was designed as an exact replication of the monkey study.

METHOD

Subjects.—Sixteen male and 16 female employees of the University of Wisconsin Primate Center and Primate Laboratory volunteered as *S*s. All were naive with respect to the experiment and problem.

Apparatus.—Except for the changes noted below, the apparatus described in the monkey report (Polidora & Thompson, 1964) and in an apparatus paper (Polidora & Main, 1963) was used. Briefly, it consisted of a key punch which concurrently programed stimuli and recorded responses, a test panel which contained the displays and manipulanda, and a control console. For this experiment the test panel was removed from its usual position in front of the monkey's cage and placed on a table vertically in front of *S*. The *E* adjusted *S*'s seat so that the displays were at eye level 18 in. away. Each of the two displays consisted of 12 $\frac{3}{8}$-in. square lights (elements) mounted adjacently in a 3×4 array (see Polidora & Main, 1963, Fig. 5). Instrumental responses were made by *S*'s pressing the bottom-hinged transparent panel overlaying each display. When an instrumental response was made, both displays extinguished immediately, a bell sounded once if the response was to the correct pattern, and the intertrial interval (ITI) ensued.

Trial presentation was under *S*'s control. By holding closed a spring-loaded lever switch (the analog of the monkey's mask response), *S* presented trials at a rate determined by the duration of the preset ITI and *S*'s response times. If *S*'s switch was closed, the displays were lighted at the expiration of the ITI and remained on until *S* responded; if the switch was released, the stimuli did not light, and responses to the panels were ineffective.

Stimuli.—Each stimulus pattern consisted of a luminous polygon formed by the lighting of certain elements in a display. By using the same IBM program cards as in the monkey study, the patterns presented to the human *S*s were identical to those presented to the monkeys. Briefly, only patterns composed of 5, 6, 7, or 8 elements were used. A pair of patterns was selected for each two-choice problem such that the difference between the number of "on" elements in the two patterns was either 0, 1, 2, or 3. This difference, termed Element Disparity (ED), constituted the area dimension. For example, a 5-element pattern paired with a 7-element

[1] Supported in part by grants from the United States Army Chemical R & D Laboratories and NIH Grant FR-0167. The assistance of W. B. McInvaille and L. A. Link is gratefully acknowledged.

pattern (5 × 7) produced a problem at ED 2; 6 × 7 problems were ED 1; 8 × 8 problems, ED 0; etc. All 10 combinations of the four number of elements (5 through 8) were used.

The second dimension, Contour Disparity (CD), was an index of the difference between the number of straight line segments around the periphery of each of the two patterns. Since the number of sides on any of these patterns is invariably a multiple of 2, CD was expressed as the difference in sides divided by 2, so that a minimal CD increment had a value of 1.

It should be understood that each pair of patterns had both an ED and a CD value. The 500 problems presented to each *S* constituted 10 replications of the 50 ED-CD problem classes.

Experimental design.—The experimental design was two replications of a factorial combination of 4 levels of ED (0–3), 5 levels of CD (0–4), and 5 levels of ITI (2, 4, 6, 8, and 10 sec.).

Procedure.—The only procedural differences between this and the monkey study were in the number of trials per problem and problems per session. The human *S*s had only two trials per problem, 100 problems per daily session for 5 consecutive weekdays. A 5-min. break bisected every session. Also, each session was run at one ITI, and every *S* received a different random assignment of the 5 levels of ITI to the 5 sessions. At the beginning of the first session *S* received a description of the task, an explanation of the win-stay, lose-shift strategy of solving two-trial discrimintion learning set problems, and coached practice on three standard problems.

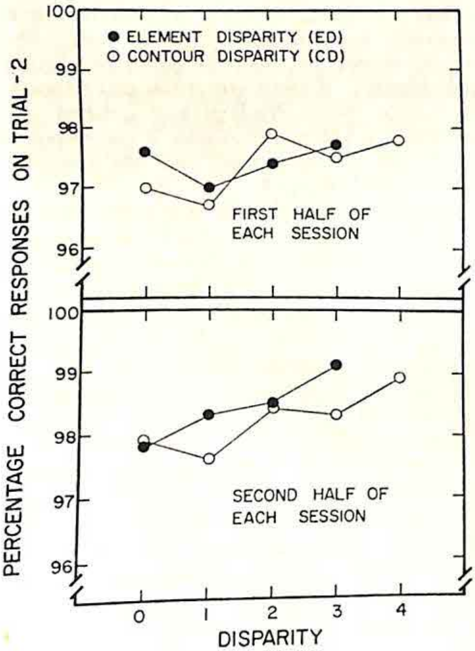


FIG. 1. Discriminative performance (percentage correct responses on Trial 2) by 32 human *S*s as a function of Element Disparity (ED) and Contour Disparity (CD). (The upper panel contains mean performance in the first half of each of the five sessions, the lower panel the second half of each session.)

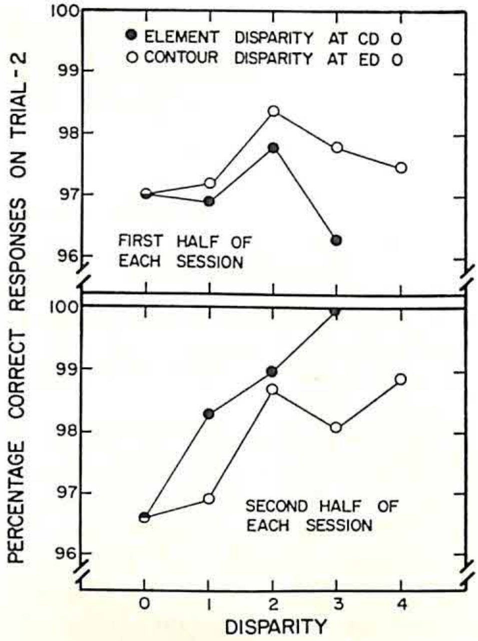


FIG. 2. Discriminative performance as a function of ED when no CD cue was available (ED at CD 0), and of CD at ED 0. (Presentation format otherwise identical to Fig. 1.)

RESULTS AND DISCUSSION

Since there was a statistically significant difference between performance on the first half and second half of each of the five sessions, $F\ (1, 30) = 13.57$, $p < .001$, the results are presented in Fig. 1 by half session. (Inspection of the data revealed that this practice or warm-up effect was most pronounced in the first session.) These results indicate that although the patterns were discriminated extremely well in general, values of ED

and CD correlated moderately well with performance within the 2% (97%–99%) range of variation. However, only the CD dimension was a statistically significant source of variance, $F(4, 120) = 4.04$; $p < .005$; ED, Sex, ITI, and all interaction sources were nonsignificant ($p > .05$).

Each point of the CD functions in Fig. 1 is the mean value computed across all levels of ED, and the converse is true for the CD functions. To determine the "pure" contributions of ED and CD to discriminability, mean performance on problems with an identical number of elements in each pattern (i.e., no area cue) was computed for each level of CD (CD function at ED 0), and the analogous function was computed for ED (ED function at CD 0). These functions, presented in Fig. 2, support an interpretation that both ED and CD were utilized as sources of discriminability.

It seems likely that a ceiling effect attenuated the performance differences along the CD and ED dimensions, differences which might otherwise have been obtained with patterns of finer "grain" (such as those used by Webster, 1963, and others cited by him) or with a measure more sensitive than percentage correct responses. Although we do not plan to study finer-grain patterns, future research will attempt to determine whether response time (latency) is as sensitive a measure of discriminability as has been suggested by Brown, Hitchcock, and Michels (1962). While these investigators have also reported essentially perfect discrimination of visual patterns by human *S*s, they found that response time in a six-choice discrimination task was related to the "complexity" (sidedness) of random shapes.

REFERENCES

Brown, D. R., Hitchcock, L., Jr., & Michels, K. M. Quantitative studies in form perception: An evaluation of the role of selected stimulus parameters in the visual discrimination performance of human subjects. *Percept. mot. Skills*, 1962, **14**, 519–529.

Polidora, V. J., & Main, W. T. Punched card programming and recording techniques employed in the automation of the WGTA. *J. exp. Anal. Behav.*, 1963, **6**, 599–603.

Polidora, V. J., & Thompson, W. J. Stimulus correlates of visual pattern discrimination by monkeys: Area and contour. *J. comp. physiol. Psychol.*, 1964, **58**, 264–269.

Webster, R. B. Pattern discrimination as a function of fill, distortion, and noise. *Percept. mot. Skills*, 1963, **16**, 361–367.

(Early publication received August 31, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 224–229

KEEPING TRACK OF SEQUENTIAL EVENTS: EFFECTS OF RATE, CATEGORIES, AND TRIAL LENGTH [1]

RICHARD A. MONTY,[2] HARVEY A. TAUB [2]
Cornell Aeronautical Laboratory, Incorporated
AND KENNETH R. LAUGHERY [3]
State University of New York, Buffalo

Ss were required to count the number of occurrences of each of several different symbols (categories) presented sequentially as a function of the number of categories (2, 3, or 4), the total number of items (8, 12, 16, or 20), and the rate of presentation (one item every 0.6, 1.0, 2.0, 3.0, and 4.0 sec.). In general, the greater the number of categories, the greater the total number of items, or the faster the rate of presentation, the poorer the performance. Of primary importance, however, is the complex interaction between these variables. The observed results are explained primarily on the basis that both presentation rate and number of categories affect the time available for rehearsal of the information while trial length results in an accumulation of errors with time.

In a series of experiments reported by Erlick (1961), *S*s were asked to identify the more frequent of two mutually exclusive events occurring in a temporal sequence. The symbols representing these events (letters and numbers) were presented one at a time as an outline of light against a dark background. In one of these experiments both the rate of presentation and the number of symbols in a sequence were varied. It was reported that at a presentation rate of one event every second, the shorter sequences resulted in better performance than the longer sequences (as measured by the increment of the more frequent event required to maintain a 75% accurate identification of that event), while at the faster rates of presentation of four and eight events per second, the reverse was true. Erlick suggests that at the slower rate, individual events could probably be discriminated one at a time and mentally tallied so that accuracy was primarily a function of memory processes. At the faster rates, however, it was suggested that the events were presented too fast to be tallied separately, and therefore *S*s reorganized the input sequences into larger units or "chunks" (after Miller, 1956) which were the basic units for the mental tally. In this case, the advantage of the longer sequence was attributed to the increased time sample which in turn allowed for a more accurate estimate of the events. Thus, it appears that two different processes were brought into play, depending on the rate of presentation of the events. At the slower rate, performance may have been primarily a function of recall or memory processes, while at the faster rates, complex perceptual processes were operating in addition to the memory processes.

[1] The work was sponsored by Cornell Aeronautical Laboratory, Incorporated, internal research funds. The authors gratefully acknowledge the assistance of Donald F. Dansereau and Charles Hall in gathering and reducing the data and the many valuable suggestions of William J. Ruby.

[2] Also affiliated with the State University of New York at Buffalo.

[3] Also affiliated with Cornell Aeronautical Laboratory, Incorporated.

The present study was designed to examine human ability to mentally tally and recall events presented sequentially in a manner similar to that of Erlick at stimulus presentation rates which clearly allowed the individual events to be discriminated. The *S*'s task in this experiment was to keep track of the number of times that each event occurred in a sequence rather than to simply indicate which event occurred most frequently. In some respects, the demands on *S* resemble those imposed by Yntema (1963). In Yntema's experiments, *S* was read a series of messages, each of which informed him that one of a number of variables (events) was in some particular state. Incoming information displaced previous information so that *S* had to essentially keep track of the present state of several variables and mentally erase past states.

Method

Subjects.—The *S*s were 150 students recruited from an introductory psychology course at the State University of New York at Buffalo.

Apparatus.—The *S* was seated in front of a panel, the face of which was positioned at a 60° angle to his horizontal line of sight. An Industrial Electronic Engineers, Incorporated in-line readout display unit was mounted in the panel. The distance from the display to *S*'s eye was approximately 2 ft. This display unit, which closely resembled the one used by Erlick (1961), contained 12 lenses, each etched with one of the symbols to be presented to *S*. By turning on and off individual lights behind each of the lenses, the symbols were projected one at a time onto approximately the same location of a $1\frac{1}{2} \times 1\frac{7}{8}$ in. frosted glass surface. The symbols, which were letters of the alphabet, were seen as an outline of light against a dark background. Six such display panels were located side by side enabling as many as six *S*s to be tested simultaneously. The *S*s were separated by partitions which effectively blocked their view of one another.

The sequences of letters flashed on the display and the rate of presentation were controlled by Grason-Stadler Company behavioral research equipment modules. This equipment was located in an adjoining room so that *S*s could not hear the noise generated by the operation of the equipment.

Procedure.—The *S*s were assigned randomly to 15 equal groups, each serving under a different combination of rate of symbol presentation and number of categories. The rates were one letter every 0.6, 1.0, 2.0, 3.0, and 4.0 sec. with the on-time to off-time ratio being approximately equal in each case. The number of categories, defined as the number of different letters appearing on each trial, were 2, 3, and 4. When two categories were presented, the letters were Q and R; for three categories, Q, R, and S; and for four, Q, R, S, and T.

Trial length, defined as the total number of letters presented on a trial irrespective of the category to which they belonged, was included as a within-*S*s variable. All *S*s were presented with trial lengths of 8, 12, 16, and 20 letters. To permit the assessment of practice, a total of 32 trials was presented. Each of the four trial lengths was presented twice in random order during the first 8 trials and twice more in different random orders in each of the remaining three blocks of 8 trials. There were four different sequences of letters for each of the four different trial lengths so that *S*s were exposed to each of the different sequences only twice: once in the first 16 trials and once in the second 16 trials. The sequences of letters were chosen at random with the restriction that each category appeared at least once on every trial.

All *S*s were instructed to mentally count the number of times that each of the different categories (letters) was presented in a trial and to record the answer at the end of the trial. Two practice trials were given to be sure that each *S* understood the instructions. These trials were followed by 32 experimental trials with a 15-sec. intertrial interval for recording answers. A loud clock signaled the beginning and end of each trial. In all cases, *S*s had knowledge of the number of different categories to be presented and the specific letters that would appear, but no knowledge of the total number of letters to be presented per trial. Finally, *S*s received no instructions concerning any method which might be utilized to mentally count the number of times each category appeared.

Upon completion of the data collection, *S*s were asked to describe in writing the techniques used to keep track of the number of occurrences of each category.

RESULTS

Three indexes of performance were examined: (*a*) absolute error, defined as the sum of the absolute differences between the number of letters presented in each category and the number reported in each category; e.g., if 5 Qs and 3 Rs were presented and *S* reported 6 Qs and 2 Rs, the number of errors for that trial would be a total of 2, one for the Q category and one for the R category, (*b*) percentage error, defined as the ratio of the absolute error per trial to the trial length multiplied by 100, and (*c*) *S*s' written comments.

The absolute error served as the primary measure of performance on each trial for each *S*. To assess the effects of practice, the two scores of equal trial length within each block were summed. This resulted in a single score for each of the four trial lengths within each of the four blocks. These scores were subjected to an analysis of variance with Rate of Presentation and Categories as between-*S*s variables and Trial Length and Blocks as within-*S*s variables. The data contributing to the four significant main effects revealed that performance varied inversely with number of categories, $F(2, 135) = 52.9$ ($p < .001$), trial length, $F(3, 405) = 200.9$ ($p < .001$), and rate of presentation, $F(4, 135) = 48.3$ ($p < .001$) and varied directly with amount of practice, $F(3, 405) = 18.1$ ($p < .001$).

Of greater importance, however, are the various interactions between these variables. The data underlying the significant Trial Length × Categories interaction, $F(6, 405) = 14.3$ ($p < .001$), suggests that, as the length of the trial increased, performance degraded in a linear fashion and that the rate of degradation increased as the number of categories increased. Similarly, the data for the Trial Length × Rate interaction, $F(12, 405) = 18.4$ ($p < .001$), indicates that, at the slowest rate of presentation, (one letter every 4 sec.), there were only slight differences in performance with the various trial lengths, but as the rate increased, these differences became gradually more pronounced. The data underlying the Trial Length × Rate × Categories interaction which yielded an $F(24, 405) = 2.0$ ($p < .01$) have been plotted in Fig. 1. It can be seen that when two categories of information were presented, mean error increased as trial length increased at the faster stimulus presentation rates of one letter every 0.6 sec. and 1 sec., respectively. However, at the slower rates of presentation there was little change in performance with an increase in trial length except for a slight increase in error with a trial length of 20 letters. In other words, when the rate of presentation was one letter every 2 sec. or less, *S*s kept track of events with almost perfect accuracy regardless of the trial length. When three categories were presented, increases in mean error as a function of trial length for the 0.6- and 1-sec. rates were even more pronounced. In addition, the 2-sec. rate resulted in a similar increase in error although of less magnitude. Finally, with four categories of information, all rates of presentation resulted in an increase in error as a function of an increase in trial length. The mean error at a rate of one letter every 0.6 sec. resembled the mean error at the 1-sec. rate. Similarly, at rates of one letter every 2 sec. and every 3 sec., performance was nearly identical, yet at the same time, decidedly better than performance at the two faster rates. Finally, performance at the 4-sec. rate was

superior to performance at any of the above rates.

The interactions of Blocks × Categories, F (6, 405) = 2.2 ($p < .05$), Blocks × Trial Length, F (9, 1215) = 5.9 ($p < .001$), and Blocks × Trial Length × Categories, F (18, 1215) = 2.2 ($p < .01$), suggest that differential practice effects were present. In general, the data indicate that the amount of improvement over blocks varied directly with length of trial and number of categories. When two categories were presented, there was no evidence of learning. With three categories, learning occurred only for the trial lengths composed of 16 and 20 letters. Finally, with four categories, performance improved across the first three blocks for the 12-, 16-, and 20-letter trial lengths, and across the last two blocks for trial lengths of 8 letters. Although learning was dependent upon both trial length and number of categories, the significant main effects for Categories and Trial Length were in evidence even on Block 4. The fact that all the interactions with both Blocks and Rate as components failed to reach significance indicates that improvements in performance attributed to practice were not related to the rate of presentation of the stimuli.

With respect to percentage error, the second performance index, it was found that, in general, for a given rate of presentation and a given number of categories, performance across trial lengths was essentially invariant. To be more specific, as trial length increased from 8 to 20, there was only a 3% increase in mean percentage error with two categories, a 2% increase with three categories, and a 5% increase with four categories. In short, had percentage error been used as the primary performance measure, the interactions with Trial Length as a component would have failed to reach statistical significance.

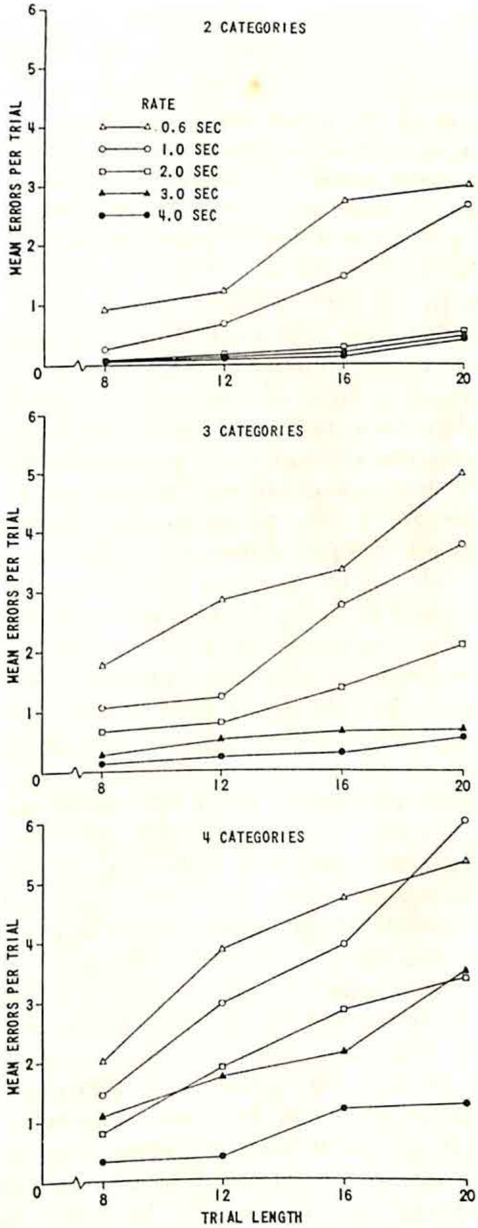


FIG. 1. Mean error per trial as a function of categories, rate, and trial length.

In their written comments, most of the Ss indicated that during early trials they had tried several different techniques of keeping track of the number of occurrences of each cate-

gory but eventually settled down to one they considered effective. With the exception of those tested at the rate of one event every 0.6 sec., 72% of the *S*s indicated that they found it most useful to mentally picture a set of windows, one corresponding to each of the different letters presented. Individual running tallys were then kept in each window and only the tallys were rehearsed after each presentation of a new event. By contrast, another 16% of these *S*s indicated that they rehearsed both the letter and the tally after each presentation without the aid of any form of mental picture. For example, the former group simply rehearsed 5, 4, 2, 3, while the latter would have rehearsed 5 Qs, 4 Rs, 2 Ss, 3 Ts. The remaining 12% reported using a variety of techniques including physical aids such as tallying on their fingers. At the presentation rate of one letter every 0.6 sec., the picture was somewhat different. With two categories, 7 of the 10 *S*s indicated that they actually rehearsed only one of two categories each time and made an estimate of the number of occurrences of the other at the end of the trial. A similar report was made by 7 of the 10 *S*s presented with three categories. In this case again, only one category was actually tallied and rehearsed, while the other two were estimated. Finally, with four categories at a rate of one letter every 0.6 sec., there was almost no agreement between *S*s regarding the technique employed. Further, there was some indication that many of the *S*s were still searching about for an effective technique at the end of the 32 trials.

Discussion

In Erlick's (1961) experiment on rate of presentation, *S*s were asked to indicate which letter appeared most often rather than to tally the actual number of occurrences of each letter. He expressed performance in terms of the increment of the more frequent event over the less frequent event required to maintain a 75% accurate identification of that event. Thus, differences in both *S*s' task and the performance measures employed preclude a direct comparison of performance at those rates of presentation common to both studies. However, the small number of errors obtained with two categories in the present study, particularly at the shorter trial lengths lends support to his hypothesis that, at slow rates of presentation (i.e., one event per second or less), individual events falling into two categories can be discriminated one at a time and mentally tallied. Erlick further suggests that at these slow rates if the events can be discriminated and tallied one at a time, then it is memory processes which play an important role in the degree of accuracy. Verbal reports of *S*s in the present study add credence to this hypothesis insofar as they indicate the extent to which rehearsal techniques were employed. It appears that, as long as the presentation rate is slow enough to allow for rehearsal of all categories (assuming the number of categories is small and *S*'s memory span is not exceeded), differences in performance are not likely to result as either the number of categories or the trial length increases. However, if the number of categories is increased to the point that a given rate of presentation precludes complete rehearsal of all items, then the number of errors made should be a function of the trial length. That is, if the last state of the items in each category has not been rehearsed at the time another symbol is perceived, an error is more likely to occur in remembering the states of the unrehearsed categories which in turn may result in an error in the new total or tally. Once an error is made, it is not likely to be corrected at a later time, particularly if the error itself is rehearsed. Errors then should be cumulative, and the longer the trial length, the greater the number of errors expected. It is therefore possible to account for the

observed interaction between Rate, Categories, and Trial Length primarily on the basis of time available for rehearsal. However, it is not reasonable to generalize this explanation to situations involving a greater number of categories, because other factors may be involved such as those associated with the classic digit span. Further research is necessary to determine the effect of increasing the number of categories beyond four.

Although the present study utilized a less complex set of stimuli and required an incremental tally as compared to the substitution of new messages for old required in Yntema's (1963) study, there is a striking similarity in the general conclusion of both investigations. Both studies indicate that for a given trial length it is better to present few categories with many possible occurrences of each than to present a large number of categories with fewer occurrences of each. The similarity of this finding suggests that rehearsal may also be critical in the task used by Yntema.

Assuming that task difficulty was increased by increasing the trial length, the number of categories presented, and the rate of presentation, it is not surprising to find differential practice effects as a function of the first two parameters—generally speaking, the more difficult a task, the greater the improvement in performance with practice. It is somewhat surprising, however, that there were no differential practice effects associated with rate of presentation. A possible explanation might be that an insufficient number of trials were run for practice to begin to have a differential effect on performance as a function of rate. As indicated above, questioning of *S*s after testing revealed that the large majority cast about for a counting technique and tried several before they found what they considered to be a satisfactory one. It is possible that at the faster rates where the greatest gain might be expected, *S*s spent more trials searching about for a technique than they did at slower rates and thus delayed the benefits to be derived from practicing with a specific technique. This possibility should be the subject of further research.

REFERENCES

ERLICK, D. E. Judgments of the relative frequency of a sequential series of two events. *J. exp. Psychol.*, 1961, **62**, 105–112.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

YNTEMA, D. B. Keeping track of several things at once. *Hum. Factors*, 1963, **5**, 7–17.

(Early publication received November 4, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 230–236

EFFECT OF IRRELEVANT INFORMATION ON A COMPLEX AUDITORY-DISCRIMINATION TASK [1]

WILLIAM E. MONTAGUE [2]

University of Illinois

An attempt was made to demonstrate that the detrimental effects of irrelevant information on discrimination learning are due to implicit response competition generated by task conditions. In a complex auditory-discrimination task, groups receiving different amounts of irrelevant information (1, 2, or 3 dimensions) never relevant to their task made fewer errors than groups receiving different amounts of irrelevant information which sometimes required differential responding. In addition, the errors increased with the number of irrelevant dimensions. Variation of task difficulty by manipulating the discriminability of the relevant information resulted in an enhancement of the effects of the irrelevant information. Repeated practice sessions reduced the effects of irrelevant information which was sometimes relevant at a faster rate than for the irrelevant information which was never relevant to the task.

In many situations humans are faced with discrimination tasks made difficult by the presence of extraneous or irrelevant factors. Studies of tasks in which signals must be detected or identified from among others of a similar nature show performance decrements brought about by irrelevant factors, e.g., Pollack (1955); Poulton (1953); Peters (1954); Pollack and Klemmer (1954); Egan, Carterette, and Thwing (1954); Webster and Thompson (1954). Although these studies show that the presence of information irrelevant to the task seriously degrades performance, little attempt was made to specify the locus of the interference. In addition, certain seemingly contradictory evidence exists. For example, Archer (1954) found that the presence and amount of irrelevant information had little, if any, effect on the proficiency of identifying relevant dimensions in a complex visual discrimination task. In experiments involving speech communication Cherry (1953), Webster and Solomon (1955), and Spieth and Webster (1955) report that under certain conditions irrelevant information interferes with the reception of relevant words while under other conditions it has less or no effect. This implies that there are degrees of interference by irrelevant information, the level perhaps being determined by aspects of the task. Archer (1962) indicated that the discriminability or obviousness of irrelevant information plays an important part in determining the effects of irrelevant information. Broadbent (1958) makes a similar point regarding selective attention. When irrelevant dimensions are obvious or attention-getting they elicit responses which interfere with performance in comparison to situations where they are less obvious. When the relevant and irrelevant dimensions are equally obvious con-

[1] The author wishes to thank John C. Webster and Milton H. Hodge for their thoughtful comments and assistance.

[2] Also at the United States Navy Electronics Laboratory, San Diego, California where this research was conducted.

The opinions and assertions contained herein are the private ones of the writer, and are not to be construed as official, or as reflecting the views of the Navy Department or the Naval service at large.

cept identification is difficult (Archer, 1962). The level of interference should also be a function of the similarity between relevant and irrelevant dimensions. Peters (1954), Speith and Webster (1955), and Broadbent (1958), among others, report that as the irrelevant information is made physically dissimilar to the relevant information, e.g., by filtering irrelevant vocal messages, there is little performance decrement. Therefore, it seems reasonable to assume that implicit response competition produces the effects of irrelevant information and that when conditions obtain which provide for increasing amounts of response competition, serious performance decrements in the form of errors and/or response speed should be observed.

Hodge (1959) provided a direct test of this assumption. In his task, context within a trial determined relevancy, that is to say, dimensions irrelevant to a particular discrimination trial were relevant on others providing conditions for high levels of response competition and poor performance. Irrelevant dimensions present on a trial which were relevant on other trials interfered with performance more than irrelevant dimensions which were never relevant to the task. The present experiment was designed as an extension of Hodge's using auditory stimuli. Briefly stated, its purpose is to demonstrate that in making complex discriminations, when conditions are provided for increasing amounts of response competition, performance is progressively degraded. Two conditions of relevancy were used to produce different levels of response competition, i.e., conditions where irrelevant dimensions which were sometimes relevant to the task were contrasted with conditions where the irrelevant dimensions were never relevant to the task. Since more interference should be produced by greater amounts of irrelevant information, one, two, or three irrelevant dimensions were used. In addition, in order to determine whether the effects of kind and amount of irrelevancy are enhanced by task difficulty, the difficulty of discrimination was varied. Three practice sessions were given to observe any systematic changes in the effects of the other variables with increasing experience.

Method

Subjects.—The *S*s were 42 undergraduate students attending San Diego State College. They were paid for their participation.

Apparatus.—The apparatus consisted of a tape playback unit, with earphones and a set of response switches for *S*, and the appropriate recording devices. The stimulus material was received from the tape recorder at *S*s' PDR-8 earphones at about 80 db. Four response switches were mounted in a small (10 × 4 × 2 in.) box and could be moved in two directions. The switches were spring loaded and returned to neutral (central) position when released. The *S*s held their hands on the response board at all times. When *S* moved a switch away from its central or rest position a corresponding indicator light lit on *E*'s control panel. Stop clocks calibrated in .01 sec. measured response latencies. The clocks were activated at stimulus onset and stopped when *S* made his response. Three sets of earphones and response boxes were provided so that three *S*s could be run at one time.

Stimulus material.—The stimuli were 4-sec. samples of a combination of dimensions obtained using a white noise and pure tones. As many as six bivalued dimensions were encoded in a sample. Two dimensions were present in all samples while the occurrence of the older four depended upon the experimental conditions holding during a particular sequence. These will be referred to as primary and secondary dimensions, respectively.

The basic signal was a white noise which was modulated slowly, yielding a rhythmic sound waxing and waning about once every second. Some dimensions were obtained by varying this signal, others were obtained by the addition of other signals. The two primary dimensions and their discriminable values were as follows: (*a*) *Pitch.* The white noise was passed through a $\frac{1}{3}$-octave filter and

TABLE 1

EXAMPLES OF THE LABELING FOR THE RESPONSE BOARD FOR THE DIFFERENT IRRELEVANCY GROUPS

Sometimes Relevant Group				Never Relevant Group			
High Pitch		Low Pitch		High Pitch		Low Pitch	
Loud	Soft	Soft	Loud	Loud	Soft	Soft	Loud
Fast Beat *[a]	High Warble *[a]	Accent 2 *[a]	Sweep Up *[a]	Fast Beat *[a]	Fast Beat *[a]	Fast Beat *[a]	Fast Beat *[a]
Slow Beat	Low Warble	Accent 4	Sweep Down	Slow Beat	Slow Beat	Slow Beat	Slow Beat

[a] * indicates switch position.

two values were chosen and designated as *high* and *low* pitch. The center frequencies for the high and low pitches were set at 4,000 cps and 800 cps, respectively. (*b*) *Loudness.* Two levels were designated as *loud* and *soft.* Since loudness and pitch interact, a psychophysical study was run with eight *S*s in order to obtain samples that sounded equally loud. The four secondary dimensions and their values were: (*a*) *Beat speed.* The rate of modulation of the white noise was varied. Two different rates were called *fast* (6/sec) and *slow* (3/sec). (*b*) *Accent.* Certain of the beats were accented or emphasized. That is to say, their level was raised several decibels above the other beats. The judgment required concerned which beat was accented. Either every *second* or every *fourth* beat was accented. (*c*) *Tone sweep.* During the sample a tone was swept *up* or *down* in frequency between 200 and 2,000 cps. (*d*) *Warble tone.* Warble tones of two different frequencies, *low* or *high* (500 cps and 1,500 cps) were inserted into the sample.

Stimulus samples were constructed from various combinations of the dimensions listed above and were recorded on $\frac{1}{4}$-in. magnetic tape. The two primary dimensions, pitch and loudness, were present in all samples. The use of the secondary dimensions was determined by the experimental conditions. Several sequences of 128 samples were made corresponding to variations in the task requirements and in the number of irrelevant dimensions.

In addition to the difference among the samples' dimensions listed above, the *high* and *low* pitches of the basic signal were varied systematically in order to make task difficulty a variable. A pitch continuum was created by varying the basic signal in discrete $\frac{1}{3}$-octave steps between 800 cps and 4,000 cps (center frequencies) on a Bruel and Kjaer $\frac{1}{3}$-octave filter. In this way, a pitch continuum of eight distinct pitches was formed, i.e., 800; 1,000; 1,250; 1,600; 2,000; 2,500; 3,200; 4,000 cps. The lower four $\frac{1}{3}$-octave bands were designated as low and the four high bands were designated as high pitch. Stimulus samples from the center of this continuum make difficult the identification of the sample as high or low in pitch. The psychophysical study mentioned above established levels for the different pitches so that they would sound equally loud.

Task and experimental design.—The *S*s were required to identify the stimulus samples in terms of the value of each of three dimensions present. They had to identify the value of *pitch* (high or low) and *loudness* (loud or soft) of the basic signal present in each sample. Then, *S*s identified the value of the relevant, third dimension. The movement of a switch documented *S*s' judgments about the signal. Each switch represented a unique combination of pitch, loudness, and relevant secondary dimension. The other secondary dimensions were present in different numbers for different *S*s allowing variation of the amount of irrelevancy. Task conditions were varied by having some *S*s respond to all four of the secondary dimensions on different trials during a sequence while others responded to only a single secondary dimension during the entire sequence.

Four major variables were manipulated. Task or kind of irrelevancy and amount of irrelevant information were between-*S*s variables and discrimination difficulty and practice were within-*S*s variables. Thirty-six *S*s were divided randomly into two major groups receiving the different kinds of irrele-

vant information, *sometimes* and *never relevant*. Table 1 displays examples of the response-board labeling used for the two major groups which should prove useful for understanding the differences between the tasks for the two groups. Stimulus sequences were prepared which required one group to respond to all four of the secondary dimensions an equal number of times during the sequence of trials. On a trial *S*s identified pitch, loudness, and relevant secondary dimension with the other dimensions irrelevant on that trial. However, the other dimensions were sometimes relevant in that it was necessary to identify them on other trials. During other sequences the second major group always identified the values of the two primary dimensions and one of the four secondary dimensions. The remaining dimensions were never relevant to their task and were never discriminated although they might be present in a sample. Due to the small number of *S*s used, complete counterbalancing, using all four secondary dimensions as the relevant dimension for different *S*s, was not undertaken. In two sequences either beat speed or tone sweep served as the fixed secondary dimension for the never relevant group. Half the *S*s in the never relevant condition discriminated between the two beat speeds and the other half between the tone sweep up or down. The results revealed no systematic difference between the two dimensions.

In summary, *S*s in the sometimes relevant condition identified the pitch and loudness of the basic signal and then identified the value of the relevant secondary dimension by moving the switch in the appropriate direction. The procedure was the same for *S*s in the never relevant condition except that only one secondary dimension was associated with all pitch-loudness combinations of the basic signal. These groups provide a comparison of results obtained under conditions where specific response competition is likely and where it is not.

To aid further in understanding the construction of the stimulus samples and the discriminative responses required of *S*s in the different conditions Table 2 is presented.

Equal subgroups within each major group were formed ($n = 6$) and received different amounts of irrelevant information. In the simplest condition, one irrelevant dimension was added to the relevant stimulus dimensions during a sequence, and similarly, two and three dimensions were added to the sample for the other groups. In all sequences three dimensions were relevant in a sample with either one, two, or three irrelevant dimensions present.

During all sequences discrimination difficulty was varied by introducing the 8-step pitch continuum. There were 16 samples of each $\frac{1}{3}$-octave filter setting during the 128 trial sequences given to all *S*s. All *S*s received one practice session on each of 3 days.

For comparison purposes, a control group ($n = 6$) was given an experimental sequence identical to the never relevant condition except that no irrelevant information was present.

Procedure.—The *S*s were all given instructions and preliminary training to familiarize them with their discrimination task and the response to be made. When *E* was sure that *S*s understood their task, 20 preliminary "warm-up" trials were given during which

TABLE 2

SUMMARY OF THE DIMENSIONS USED TO GENERATE THE STIMULI, THEIR VALUES, AND THE DISCRIMINATIONS REQUIRED UNDER THE DIFFERENT CONDITIONS OF IRRELEVANCY

Dimension	Values	Irrelevancy Cond.	
		Never Relevant	Sometimes Relevant
Primary		identification:	identification:
Pitch	high-low (pitch varied)	required	required
Loudness	loud-soft	required	required
Secondary			
Beat speed	fast-slow	required[a]	sometimes required
Accent	$\frac{1}{2}$–$\frac{1}{4}$	never required	sometimes required
Tone sweep	up-down	never required	sometimes required
Warble tone	high-low	never required	sometimes required

[a] Half of the *S*s in these groups were required to identify this dimension on each trial. The other *S*s never responded to this dimension and discriminated the values of warble tone instead.

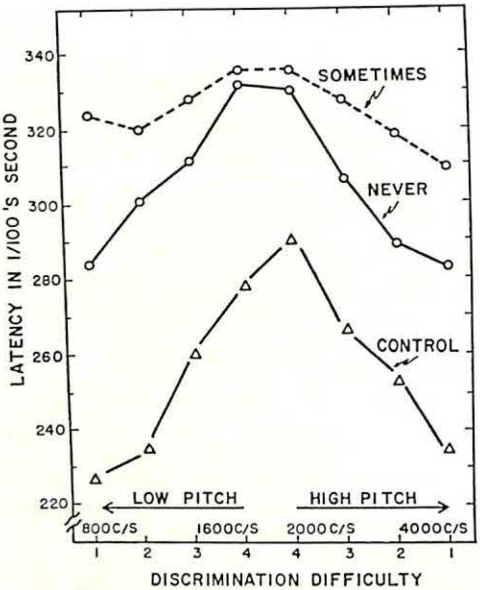


FIG. 1. Average latency as a function of position on the pitch continuum of the basic signal and kind of irrelevant information. (The central positions [4] represent the pitches most difficult and the ends [1] the easiest to discriminate. Control *S*s received no irrelevant information.)

knowledge of results was given. Then the appropriate sequence of 128 samples was given. During the experimental sequence *S*s were not told which response was correct. After each 4-sec. sample, the tape recorder was stopped until the responses and times were recorded. Then the next sample was played. Whenever possible, three *S*s in a group were run at once. On the next 2 days, after short instructions reviewing the procedure, the 20 warm-up trials were given and then the 128-trial sequence.

The response measures used were errors and response latency. The responses were recorded by *E* for scoring at a later time. Latency, the time between the onset of the sample and the response, was recorded at this time.

RESULTS

The basic data analyzed in the experiment were the 24 average latencies and error proportions made by each *S* at the eight levels of Difficulty during the 3 days of Practice. These data were subjected to analysis of variance using Kind and Amount of Irrelevant Information as between-*S*s and Difficulty and Practice as within-*S*s variables.

The analysis of the latency data revealed significant effects due to Difficulty and Practice only, $F(7, 210) = 12.645$; $F(2, 60) = 22.490$, $p < .001$ for both tests. The effects due to Kind and Amount of Irrelevant Information did not reach significance. The Kind of Irrelevant Information interacted significantly with Difficulty, $F(7, 210) = 2.580$, $p < .025$. Figure 1 shows the average latencies for the different irrelevancy conditions as a function of discrimination difficulty. Each datum point is the average of 48 samples for each pitch for each of 18 *S*s. The data for the control group were not included in the analysis and are shown for comparison purposes. Only one other interaction effect contributed enough variance to allow rejection of the null hypothesis. A significant Difficulty $\times$ Kind $\times$ Amount of Irrelevant Information triple interaction was found, $F(7, 210) = 2.038$, $p < .05$.

The analysis of the errors revealed clear-cut effects of all major variables. Figure 2 shows the average proportion

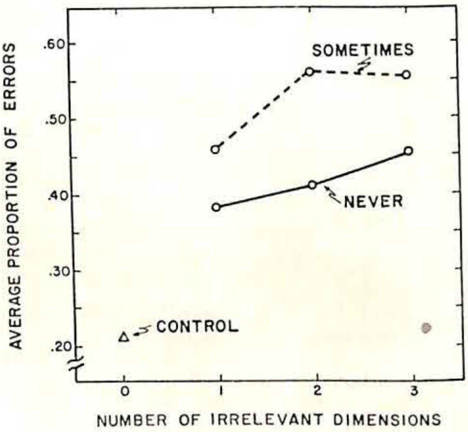


FIG. 2. Average proportion of errors as a function of the amount of irrelevant information. (Each datum point represents the average for a different group of *S*s.)

of errors as a function of the number of irrelevant dimensions for the two irrelevancy conditions. Significantly more errors were produced by irrelevant information which was sometimes relevant in comparison with that which was never relevant, $F\ (1, 30) = 30.469$, $p < .001$. In addition, increasing the number of irrelevant dimensions produces significantly more errors, $F\ (2, 30) = 6.682$, $p < .01$. The average proportion of errors made by the comparison control group is shown in the figure also. The results of Dunnett's test (Edwards, 1962) which compares all six groups with the control, reveals that all three groups receiving the sometimes relevant, irrelevant information and the group receiving three irrelevant dimensions never relevant to their task, differ significantly ($p < .05$) from the control *S*s. In the same analysis, using the average latencies, Dunnett's test revealed significant differences between any of the experimental groups and the control group. As the task of identifying the pitch of the sound was made more difficult, significant increases in errors resulted, $F\ (7,210) = 43.667$, $p < .001$. On the other hand, Practice rapidly reduced the number of errors made, $F\ (2, 60) = 58.147$, $p < .001$.

The differences between the error proportions for the two kinds of irrelevancy were not the same for all levels of Difficulty and Practice. The Kind of Irrelevancy and Difficulty interaction reached significance, $F\ (7, 210) = 2.093$, $p < .05$. Figure 3 shows the error data as a function of discrimination difficulty. Once again the data for the control group are shown.

Practice also interacts significantly with the kind of irrelevancy, $F\ (2, 60) = 4.151$, $p < .025$. The magnitude of the difference between the sometimes and never relevancy conditions decreases somewhat with practice. The only other interaction reaching significance was that between Difficulty, Practice, and Amount of Irrelevant Information, $F\ (28,420) = 1.518$, $p < .05$. However, examination of average curves for these data reveal no systematic variation of the Difficulty and Amount of Irrelevancy interaction from day to day.

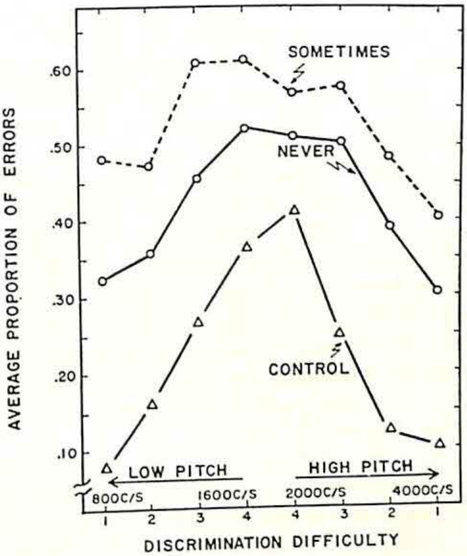


FIG. 3. Average proportion of errors as a function of position on the pitch continuum (difficulty) and the kind of irrelevant information.

DISCUSSION

The error data generally present a consistent picture of the detrimental effects of irrelevant information on complex discrimination learning. When information is irrelevant to a present discrimination but is relevant at other times it significantly degrades performance. This suggests that the locus of interference attributable to irrelevant information is in the response competition generated by implicit responses elicited by the nonrelevant dimensions. The fact that more interference results with larger

numbers of irrelevant dimensions provides further support for the hypothesis.

On the other hand, the latency data reveal no effect due to the irrelevancy conditions. This result was not unexpected, and probably was due to task conditions. The auditory-discrimination task required *S*s to judge dimensions many of which were time dependent. This would tend to attenuate any differences between groups. Support for this explanation is revealed also in the lack of any significant differences between the average latencies for the control group and the various experimental groups using Dunnett's test.

In the analysis of both errors and latencies significant Kind × Difficulty interactions were revealed. Examination of Fig. 1 and 3 reveals somewhat longer latencies and greater proportion of errors although the curves are somewhat flatter for the sometimes relevant condition in comparison with the other conditions. These results may be due to a greater amount of response competition generated both by the sometimes relevant information and the difficulty of making the initial discrimination.

The significant Practice × Kind of Irrelevancy interaction indicates that the difference between groups receiving the two kinds of irrelevancy decreases with practice. Unfortunately, three practice sessions were insufficient to determine whether this convergence continues. This possibility, however, raises an interesting question about the manner in which the effects of irrelevant information are reduced by practice. The result in Hodge's (1959) study led him to a similar concern about whether *S*s learn to ignore the irrelevant dimensions or merely learn responses to the relevant stimuli better. Ignoring stimuli which sometimes require discriminative responses would require a highly selective filtering of the informational input, which could be acquired only by responding in the presence of the competing information. If learning relevant responses mediates this convergence, practice in the absence of irrelevant information should be effective. The resolution of this problem awaits future research.

REFERENCES

Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Archer, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. *J. exp. Psychol.*, 1962, **63**, 616–620.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Cherry, E. C. Some experiments on the recognition of speech with one and with two ears. *J. Acoust. Soc. Amer.*, 1953, **25**, 975–979.

Edwards, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Holt, Rinehart, & Winston, 1962.

Egan, J. P., Carterette, E. C., & Thwing, E. Some factors affecting multichannel listening. *J. Acoust. Soc. Amer.*, 1954, **26**, 774–782.

Hodge, M. H. The influence of irrelevant information upon complex visual discrimination. *J. exp. Psychol.*, 1959, **57**, 1–5.

Peters, R. W. Competing messages: The effect of interfering messages upon the reception of primary messages. *USN Sch. Aviat. Med. res. Rep.*, 1954, No. 27. (Proj. NM 001 06 401)

Pollack, I. Visual noise filtering by human operators. I. Sequentially encoded information. *USAF Cambridge Res. Cent. tech. Rep.*, 1955, No. AFCRC-TR-54-57.

Pollack, I., & Klemmer, E. T. Visual noise filtering by the human operator: II. Linear dot patterns in noise. *USAF Cambridge Res. Cent. tech. Rep.*, 1954, No. AFCRC-TR-54-15.

Poulton, E. C. Two channel listening. *J. exp. Psychol.*, 1953, **46**, 91–96.

Spieth, W., & Webster, J. C. Listening to differentially filtered competing voice messages. *J. Acoust. Soc. Amer.*, 1955, **27**, 866–871.

Webster, J. C., & Solomon, L. N. Effects of response complexity upon listening to competing messages. *J. Acoust. Soc. Amer.*, 1955, **27**, 1199–1203.

Webster, J. C., & Thompson, P. O. Responding to both of two overlapping messages. *J. Acoust. Soc. Amer.*, 1954, **26**, 396–402.

(Early publication received November 25, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 237–240

A TEST OF THE "LIMITED CAPACITY" HYPOTHESIS [1]

BENNET B. MURDOCK, JR.[2]

University of Vermont

If as suggested recently an individual does have a limited capacity for recall in STM, then simultaneous variation in number of presentations and rate of presentation (with total presentation time constant) should result in equal recall. To test this hypothesis, Ss were given 1 presentation of 6 pairs at a rate of 4 sec/pair, 2 presentations at 2 sec/pair, or 4 presentations at 1 sec/pair, then tested for recall of the B member of 1 of the 6 A-B pairs. Recall probabilities for the 1st and 3rd condition were indistinguishable, while the 2nd condition was superior only at the shorter retention intervals. Considering this and other evidence, it was concluded that the limited capacity hypothesis may be a useful 1st approximation.

Recently it was suggested (Murdock, 1964) that, in STM, individuals may have a limited capacity for immediate recall. Further, it was suggested that the limitation may be on the rate of processing information. If so, then one should be able to trade off number of presentations and presentation rate if total presentation time is held constant. For instance, if each pair in a list of paired associates is presented for a total of 4 sec., the list can be presented once at a presentation rate of 4 sec/pair, twice at a rate of 2 sec/pair, or four times at a rate of 1 sec/pair (symbolized 1-4, 2-2, and 4-1, respectively). The hypothesis would predict equal recall under these three conditions; the present experiment is a test of this hypothesis.

METHOD

To facilitate comparison with other studies the procedure was essentially that used previously (e.g., Murdock, 1963a, 1963b). All A-B pairs were composed of common English words paired at random, and each list consisted of six pairs. After 1-4, 2-2, or 4-1 one of the six pairs was tested by presenting A as the cue for recall of B. Each of the six serial positions was tested equally often under the three experimental conditions. Thus, one replication required 18 lists. In all there were 13 replications; group testing was used, and all *Ss* ($N=20$) were tested on all 234 lists. The four test sessions required were held at the same time (11:00 A.M.) on four consecutive Tuesdays.

The same filmstrips that had been used previously (Exp. III, Murdock, 1963b) were used here. For the two conditions that required repetition (i.e., 2-2 and 4-1), after the last pair in the list had been shown for the requisite time, the stripfilm was rolled back to the first pair in the list. The roll-back took no appreciable time and was just a rapid blur on the screen. This procedure was repeated twice more for the 4-1 condition. Thus, with repetition the same serial order of the pairs was preserved, and for all practical purposes there was no intertrial interval. Actually, this lack of an intertrial interval is one of the two procedural differences between the 2-2 condition of the present experiment and the two-repetition condition of a previous study (Exp. III, Murdock, 1963a). The other difference is that, in the present experiment, *Ss* were given information before each trial as to whether the presentation rate would be "fast," "slow," or "average."

Naturally, the critical pair was in no way distinguished prior to the recall test. The recall period was 15 sec., and a 5-sec. ready period preceded each list. The order of presentation of the 18 conditions within each replication was randomized. With group testing the specific A-B pairs could not be counterbalanced across the 18 conditions; however, with a total of 1,404 different word pairs randomly distributed, it seemed un-

[1] This investigation was supported by PHS Research Grant MH 03,330 from the National Institute of Mental Health, Public Health Service. Gretchen Miller and George Smith helped in the analysis of the data.

[2] Now at the University of Missouri.

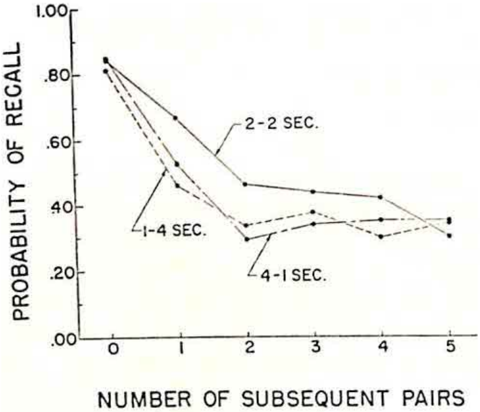


FIG. 1. Probability of recall as a function of number of subsequent pairs.

likely that any systematic differences in item difficulty could occur to bias the results. The *S*s were students of both sexes from the introductory psychology course who were fulfilling a course requirement.

RESULTS

The best overall measure of performance is the area under the serial-position curve (i.e., the sum of the p values for the six different serial positions). For the 2-2 condition the value was 3.12; the comparable value for the two-repetition condition of the previous experiment (Murdock, 1963a) was 3.47. The elimination of the 2-sec. intertrial interval reduced the total presentation time by 8%; the fact that the recall performance was reduced by 10% does not seem disproportionate.

The main results of the present experiment are shown in Fig. 1, where probability of recall is plotted as a function of number of subsequent pairs for each of the three conditions separately. It is apparent that the results for 1-4 and 4-1 are almost indistinguishable; however, the p values for 2-2 are somewhat higher at the shorter retention intervals (in particular, one and two subsequent pairs).

The above conclusions are substantiated by analyses of variance of number of correct recalls. Number of subsequent pairs, experimental conditions, and their interaction were all significant at well beyond the .001 level. Duncan's multiple-range test showed that, at the .01 level, 2-2 was better than either 1-4 or 4-1 but the latter two did not differ significantly. Separate analyses of variance at each of the six retention intervals showed that the experimental conditions were significant only for one and two subsequent pairs; at both these retention intervals $p < .001$. Finally, it has been suggested (Murdock, 1963a) that, after two subsequent pairs, the retention curve is essentially asymptotic. Therefore, the three longest retention intervals (i.e., three, four, and five subsequent pairs) were analyzed separately; an analysis of variance showed that the effects of neither subsequent pairs nor experimental conditions were statistically significant (in both cases, $p > .05$).

As stated above, the area under the serial-position curve was 3.12 for the 2-2 condition. The comparable values for 1-4 and for 4-1 were 2.62 and 2.70, respectively. Thus, in terms of overall performance, the latter two conditions were about 15% poorer than the 2-2 condition. However, as indicated by the statistical analysis, most of the difference appeared to be

TABLE 1

PROPORTION OF OMISSIONS, EXTRALIST INTRUSIONS, AND INTRALIST A AND B INTRUSIONS FOR EACH EXPERIMENTAL CONDITION

	1-4	2-2	4-1
Omissions	.558	.487	.407
Extralist intrusions	.200	.179	.127
Intralist A intrusions	.079	.085	.113
Intralist B intrusions	.163	.250	.353

concentrated at one and two subsequent pairs.

There was a total of 2,496 noncorrect responses, and an analysis of errors is shown in Table 1. This table shows the proportion of omissions (failure to respond), extralist intrusions (any word not in the list), and intralist intrusions (any incorrect word from the list) broken down into A and B items. As number of presentations increased from one to four, both omissions and extralist intrusions decreased but intralist intrusions (both A and B) increased. Perhaps the most important point to note about the error analysis is that, for all four categories, the trend was monotonic; in no case was there a maximum or a minimum for the 2-2 condition.

To test for practice effects the first three replications were compared with the second three replications. An analysis of variance of the number of correct recalls showed that neither replications (i.e., 1–3 vs. 4–6) nor the interaction of replications by conditions (i.e., 1-4, 2-2, 4-1) had a statistically significant effect on recall (in both cases $p > .05$). An inspection of the data gave no indication that learning to learn and proactive inhibition were counteracting each other. Performance at Serial Position 6 (probably the best measure of degree of original learning) did not appear to improve over these first six replications; neither did performance at Serial Positions 1–3 deteriorate over this period. However, such changes may have occurred within the first replication itself (see Murdock, 1964).

Finally, as the number of presentations increased from one to four, the opportunity for serial learning of the six pairs increased. (The same serial order was used with repetition.) One measure of such serial learning is the "adjacent/penultimate" (A/P) ratio. In previous studies we have found that about two-thirds of the intralist intrusions are either from adjacent serial positions (i.e., immediately before or immediately after the critical pair) or from the penultimate pair in the list (i.e., Serial Position 5). With serial learning of the pairs the former could be expected to increase relative to the latter (see Slamecka, 1964) and, in one study (Exp. III, Murdock, 1963a), this ratio was 1.08, 1.38, and 1.61 for one, two, and three presentations, respectively. In the present study the A/P ratio was 1.48, 1.41, and 1.63 for one, two, and four presentations, respectively. Thus, it would seem as if relatively little serial learning occurred; presumably the faster presentation rate worked against the factor of repetition.

Discussion

In general the "limited capacity" hypothesis would seem to be confirmed. For the longer retention intervals (i.e., 3–5 subsequent pairs) the three conditions did not differ significantly, and at the shorter retention intervals (i.e., 1–2 subsequent pairs) 1-4 and 4-1 did not differ significantly. Also, the trend of omissions, extralist intrusions, and intralist A and B intrusions was monotonic; thus, this measure of retention failed to show any evidence of curvilinearity across conditions. The one disquieting feature was the superiority of 2-2 at the shorter retention intervals. However, there is some evidence already that there are two components to the STM paired-associate retention curve (Murdock, 1963a), yet even here the differences that did exist were only of the order of 15%.

Recent studies by Bugelski (1962) and by Bugelski and Rickwood (1963) are particularly relevant to the present findings. In the former study a list of paired associates was learned to criterion at one of five different presentation rates. While the number of trials to criterion

increased with presentation rate, the total exposure time for the five groups was essentially the same. In the latter study the trials were self-paced, yet the total learning time was clearly within the range established in the first study. Obviously, these results are just what would be expected by a hypothesis of limited capacity. Additional support comes from Nodine (1963) who reported that, ". . . an eightfold increase in exposure durations . . . produced slightly less than an eightfold increase in total correct responses [p. 105]."

There is some evidence, then, that with paired associates a constant amount of material can be learned or retained given a constant amount of time. The same time constancy seems to operate in free recall (Cohen, 1963; Murdock, 1960; Waugh, 1962, 1963). If the limitation is on the rate of processing information, then these studies in verbal learning would agree well with studies of choice reaction time (Leonard, 1961). All things considered, it would seem that the limited capacity hypothesis may be a useful first approximation.

REFERENCES

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 409–412.

BUGELSKI, B. R., & RICKWOOD, J. Presentation time, total time, and mediation in paired-associate learning: Self-pacing. *J. exp. Psychol.*, 1963, **65**, 616–617.

COHEN, B. H. An investigation of recoding in free recall. *J. exp. Psychol.*, 1963, **65**, 368–376.

LEONARD, J. A. Choice reaction time experiments and information theory. In Colin Cherry (Ed.), *Fourth London symposium on information theory.* Washington, D. C.: Butterworth, 1961.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

MURDOCK, B. B., JR. Short-term memory and paired-associate learning. *J. verbal Learn. verbal Behav.*, 1963, **2**, 320–328. (a)

MURDOCK, B. B., JR. Short-term retention of single paired associates. *J. exp. Psychol.*, 1963, **65**, 433–443. (b)

MURDOCK, B. B., Jr. Proactive inhibition in short-term memory. *J. exp. Psychol.*, 1964, **68**, 184–189.

NODINE, C. F. Stimulus durations and stimulus characteristics in paired-associate learning. *J. exp. Psychol.*, 1963, **66**, 100–106.

SLAMECKA, N. J. An inquiry into the doctrine of remote associations. *Psychol. Rev.*, 1964, **71**, 61–76.

WAUGH, N. C. The effect of intralist repetition on free recall. *J. verbal Learn. verbal Behav.*, 1962, **1**, 95–99.

WAUGH, N. C. Immediate memory as a function of repetition. *J. verbal Learn. verbal Behav.*, 1963, **2**, 107–112.

(Received January 4, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 241–245

SIMILARITY IN STIMULUS MATERIAL AND STIMULUS TASK ON THE FORMATION OF A NEW SCALE OF JUDGMENT

M. E. TRESSELT

New York University

180 Ss were tested individually, 36 Ss in each of 5 groups, in order to examine the effects of 2 different prior set task situations and 2 different prior set anchoring stimulus materials on the formation of a new scale of judgment. The control group was given a prearranged order of a test series of 12 weighted cardboard containers (11–560 gm.) and asked to judge each weight as heavy, medium, or light. 2 of the experimental groups were asked 1st to make judgments of a series of books (612–1,184 gm.) ranking them in heaviness or to make judgments on a series of plastic weighted containers (612–1,184 gm.) before being given the test series of weights. 2 other experimental groups were asked 1st to alphabetize books or alphabetize the plastic weighted containers as an irrelevant task before being given the prearranged 12 weighted containers. The results confirm earlier research that the control group shows a gradual adjustment of the scale of judgment to the objective scale of weights over a period of exposure. Also the results show that those Ss having handled the plastic weighted containers 1st showed greater displacement of the new scale than the groups exposed to books. Further, the task situation, i.e., weighing or alphabetizing, failed to produce any significant displacement.

Whenever an *S* is faced with a series of stimuli and required to make judgments along a given continuum, some anchoring stimulus or stimuli usually are used as referents. The anchor value may be one that is present, i.e., a 12-in. ruler for length; or an anchor that is not present but which is in the past experience of *S*, e.g., a memory of how long the longest line he has seen may have been, or perhaps, the standard referent can be in the form of an instruction such as the instruction to consider a yard as a long line, a foot as a short line.

When *S* is faced with a situation in which he is aked to make judgments of a series of weights, for example, to scale them as heavy, medium, and light, he would call from his memory as representative anchoring weights referents of similar properties to the stimuli with which he is faced, and make comparisons. Rogers (1941) using the method of single stimuli, called for judgments in six absolute categories with and without an experimental anchoring stimulus being present, and so demonstrated assimilation to various degrees of displacement of the anchoring stimulus from the critical series of weights. Tresselt (1947) conducted an experiment in which practice was given on a range of weights before *S*s were given an expanded range and found that the judgments came to agreement with the new expanded scale. A weight, then, that was perceived as "light" in the practice scale and now was a middle weight of an expanded scale, became "medium" as *S* experienced the new stimulus range.

In order to demonstrate the effect of similar and nonsimilar materials as referents, Brown (1953) used weights and metal trays as anchor materials. He found the greater effect on a

subjective scale of weights was produced by the anchor weights (similar materials) as compared to the tray anchor weights (dissimilar materials). He also included anchor weights where no judgment of them was required and found these more effective on the critical series than tray anchors not judged. The anchor weights not judged were less effective in influencing judgments than the judged anchor-weight condition, and the nonjudged tray anchors were the least effective of all his conditions.

The effect of nonrelevant aspects in the judgmental situation has been studied most recently by Rambo (1962) who presented his *S*s with black dots on white cards where both dot size and card size varied in different relationships, i.e., random, positively related, negatively related, etc. He found that incidental exposure of stimuli acted as anchors for the intentional judgments made later to a restricted distribution of the same stimuli.

In these latter two experiments, *S* is presented with material which he is not to judge or about which he is not required to make a verbal report in the anchoring conditions. The question may be raised that *S*s may well have been suspicious of the materials and although nonverbalized, they may have utilized the scale referents because of the artificiality of a laboratory situation. Johnson (1944) points out that in his experiments, *S*s would often move a chair or lift a book during their rest pauses without affecting their scale of values which were based on lifting stimulus weights.

There are two problems, then, to be investigated here. The first problem relates to the similarity of materials constituting both the anchoring and test series of stimuli. It is predicted that the scale of judgment of a group of *S*s given anchoring materials similar to the test series, i.e., weighted containers, will show a greater displacement in the formation of their new scale than a group of *S*s given anchoring stimuli of dissimilar materials, i.e., books. The second problem examines the part played by degree of similarity in task in the formation of a new scale of judgment. Tresselt (1948) using professional weight lifters and watchmakers as *S*s and using a series of 12 weighted containers as stimulus materials found that the center of the scale of the weight lifters was significantly different from the scale of the watchmakers. The hypothesis was suggested that the more similar the previous task to the task at hand, the more apparent would be the effect of the old scale on the adjustment to a new scale. It is predicted, therefore, that *S*s required to make judgments of a similar nature to both the anchoring stimuli and the test stimuli, i.e., weigh them, will be more affected by the anchoring stimuli in forming a new scale than a group of *S*s required to do a dissimilar task, i.e., alphabetize the anchoring series and to weight the test series. All of the experimental groups should show more of an effect on the center of the scale than a control group having a heterogeneity of previous experiences prior to the test series of stimuli.

Method

Apparatus and Procedure

There were two kinds of anchoring stimuli: books and weights. The five books alphabetically by author and by weight were in an increasing order, i.e., one by Allport weighed 612 gm.; Bills, weighed 841 gm.; Crane, 927 gm.; Guilford, 1,138 gm.; and Patterson, 1,184 gm. The other set of anchoring stimuli were five weighted black plastic containers (larger in size than the test stimuli) labeled All (612 gm.); Bil (841 gm.); Cra

(927 gm.); Guil (1,138 gm.); and Pat (1,184 gm.)

The test stimuli were 12 weighted cylindrical cardboard containers whose weight in grams were 11, 60, 110, 160, 210, 260, 310, 360, 410, 460, 510, and 560.

There were five groups of *S*s, 36 *S*s in each group, a total of 180 *S*s tested individually. Group A (the control group) was given only the 12 test weights in a prearranged order and asked to judge the weights as heavy, medium, or light. The order of the test weights for the experiment as a whole was such that each weight was presented in a different serial position to each *S*. Thus the 12 weights were lifted with equal frequency in each of the 12 serial positions. In this way, for example, the 11-gm. weight was in the first position for three *S*s in any group; in the second position for three more, the third for each of three more, and so on through the twelfth position. All *S*s were blindfolded during the test series.

The *S*s in Group B were told on entering the office for testing that by accident some books were left right in the middle of the testing table either because *E* was trying to straighten the office or because someone had brought the books back during the time *E* was absent. The *S* was asked to give the books to *E* in alphabetical order one at a time so that they could be placed into the bookcase. The *S*s were then told to sit down to do the experiment which was administered in the same way as for Group A.

Group C was informed upon entering the office that someone previously had dropped one of the plastic containers of weights (a broken one was on the desk) and since she was not going to use this particular series any more *S*s were asked to give her the weights in alphabetical order (as labeled) so that she could put them away until *E* could get a new weight to replace the broken one. The *S* was given the extra caution not to break or to drop any of them; thus to hand them to *E* one at a time. The *S* was then given the test stimuli.

Group D was told that there "seemed to be a variable present in the few subjects already tested because some people seemed to be weight-sensitive and others not." Since "there was no test for this ability," *E* had just taken five books arbitrarily from her shelf. The *S*s were asked to weigh the books, presented haphazardly on the desk, and arrange them in order from the lightest to the heaviest. The *S*s were then given the test stimuli as the "experiment." Group E was given the same instructions but using the plastic containers as the so-called weight-sensitive pretest followed by the "experiment."

Since the actual weight in grams and alphabetical order were identical for all anchoring materials, each *S* had the anchoring experiences in the identical serial order. All *S*s in the experimental groups were asked whether they had associated the anchoring task and the "experiment" together after the session was completed. Only three verbalized any insight into the problem and their data were eliminated and new *S*s added in their places.

Results and Discussion

The mean weight judged "medium," the frequencies of these judgments, and their standard deviations are presented in Table 1.

Because of the nature of the experimental plan, the Friedman two-way analysis of variance by ranks was used. A $\chi_r^2 = 16.78$ was obtained which is significant at the .01 level. The most effective anchoring seems to be reflected in judgments by those individuals who alphabetized the weights (Group C), next by those who weighed the weights (Group B); the group who alphabetized the books (Group B); those who weighed the books (Group D), and finally the control group (Group A) having no anchor had the lowest general mean value for the weights perceived as medium over the 12 serial positions.

In order to ascertain the relative strengths of the anchoring materials on the formation of the new scale, the data of Groups C (alphabetizing weights) and E (weighing weights) were combined and compared with the results of Groups B (alphabetizing books) and D (weighing books), the prediction having been made that lifting plastic containers whether alphabetized or weighed will cause a greater change in the mean weight judged medium as compared to the judgments of those *S*s who handled books, dissimilar in shape to the test

TABLE 1

MEAN WEIGHT IN GRAMS JUDGED MEDIUM (*M*), FREQUENCY OF MEDIUM JUDGMENTS (*N*), AND *SD* AS A FUNCTION OF SERIAL POSITION FOR GROUPS A, B, C, D, AND E

Serial Position	Group A			Group B			Group C			Group D			Group E		
	M	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*
1	305.5	11	152.9	340.0	15	153.6	421.1	9	107.4	306.2	13	147.4	406.7	15	117.5
2	247.5	8	98.7	320.0	5	106.8	322.5	12	113.8	272.5	8	92.7	355.5	11	138.9
3	302.9	7	132.0	338.6	7	127.8	331.4	7	95.8	250.0	5	86.0	298.9	9	104.8
4	335.0	8	96.8	297.5	12	137.2	278.8	8	60.9	260.0	5	54.8	295.7	7	109.3
5	232.7	11	80.8	235.0	10	134.6	235.0	10	128.9	193.3	12	72.9	289.2	12	107.0
6	185.0	9	43.3	232.7	11	93.8	318.3	6	123.9	210.0	4	50.0	240.0	10	100.5
7	328.2	11	133.6	230.0	10	104.9	285.0	8	136.6	245.0	10	102.6	243.3	9	91.2
8	237.2	9	60.7	291.3	8	96.6	296.4	11	71.0	271.1	9	52.3	285.0	12	80.1
9	278.2	9	88.7	278.2	11	113.4	264.6	11	62.0	235.0	8	55.9	268.3	12	123.9
10	226.7	9	102.7	200.0	10	62.5	265.0	10	65.0	193.3	9	47.2	222.5	8	69.2
11	237.8	9	71.1	245.0	10	67.3	255.0	10	68.7	240.0	10	60.0	269.1	11	66.8
12	215.6	9	54.8	226.7	9	47.1	230.0	10	55.7	216.3	8	39.0	276.7	9	62.4

stimuli. Using the sign test, the hypothesis was supported on all 12 positions of the series. It may be said that assimilation to homogenous materials is greater than to heterogenous materials with a shift in the direction of the anchor at least over 12 trials with additional stimulus materials. This result supports the results of the judgmental sequences of weight lifters on the same critical test stimuli.

The second prediction that conscious weighing as a task would be more influential in establishing a referent than an irrelevant task or incidental task such as alphabetizing was not supported ($p = .073$). In the light of previous studies such as that of Rambo, this finding is confusing. One explanation suggests that there is an interaction effect here. Rambo used the same materials in his anchoring and test series, i.e., dots and cards, but in this experiment there is conscious weighing with and without similar materials. Since there is a greater similarity in activity for Groups D and E, there should be and there was a significant difference using the sign test ($p = .003$) in the test series in favor of larger medium judgments by Group E. When dissimilar materials are used, as in Groups B and C, it was found that there was no significant difference in the judgments of medium weights. Thus, it could be suggested that anchoring produced by similar tasks will vary with the degree of similarity of the stimulus materials such that anchoring will be most effective where both task and materials are similar. With this type of analysis, then, the statement made by Johnson on students' lifting a chair with no effect on the weight scale they had formed, is also explainable since the extraneous material was very dissimilar.

The arrangement of the stimuli in a series in which each of 12 weights was presented in each of the serial positions lends itself to another type of analysis. The greatest displacement of the subjective scale should be noted on the first trial and gradual agreement with the objective scale should occur in the later serial positions. Tresselt and Volkmann (1942) have suggested that there is an area of discontinuity between the fourth and fifth serial positions. Arbitrarily ranking the mean weight judged medium

for Positions 1–12 for each of the four experimental groups from heaviest to lightest resulted in a total average rank of 1, 2.5, 3, and 4.7 in that order for Serial Orders 1, 2, 3, and 4. After the fourth weight was lifted, the rankings followed an irregular order of 8.6, 7.5, 8.0, 4.7, 8.0, 10.6, 8.0, and 9.7. In brief, the means over the first four medium judgments followed a slight descending direction in weight and then seemed to oscillate in an irregular fashion around the objective mean medium value (285.0) of the weight series.

The results of this experiment seem to indicate that having had relatively heavy anchoring materials displaces the center of a new scale upward especially when both anchoring materials and test materials are similar in structure. When the judgments of the anchoring materials are the same or different from the judgments required on the test materials, there may or may not be a displacement of the new scale. If both task and materials are the same in both anchor and test situation, there will be a shift in the center of the scale in the direction of the anchor; but if the tasks are the same and the materials differ, there will be little or no effect. Within the limits of this experiment it seems that the effect of the anchor probably fades out on or around the fifth serial position of exposure to a new set of stimuli.

REFERENCES

Brown, D. R. Stimulus-similarity and the anchoring of subjective scales. *Amer. J. Psychol.*, 1953, **66**, 199–214.

Johnson, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, **34**, 425–436.

Rambo, W. W. Absolute judgment of a restricted distribution of context stimuli. *J. gen. Psychol.*, 1962, **66**, 169–178.

Rogers, S. The anchoring of absolute judgments. *Arch. Psychol., N. Y.*, 1941, No. 261.

Tresselt, M. E. The influence of amount of practice upon the formation of a scale of judgment. *J. exp. Psychol.*, 1947, **37**, 251–260.

Tresselt, M. E. The effect of experiences of contrasted groups upon the formation of a new scale of judgment. *J. soc. Psychol.*, 1948, **27**, 209–216.

Tresselt, M. E., & Volkmann, J. The production of uniform opinion by non-social stimulation. *J. abnorm. soc. Psychol.*, 1942, **37**, 234–243.

(Received January 6, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 246–250

CULTURAL PRIMARIES AS A SOURCE OF INTERFERENCE IN SHORT-TERM VERBAL RETENTION[1]

KENNETH A. BLICK[2]

Tulane University

Kent-Rosanoff stimulus words were separated into 3 categories having different cultural probabilities of producing R_1 (primary), R_2 (secondary), or R_{3-n} (sum of R_3 to R_n). The average cultural probabilities of R_1 for the 3 categories were .14, .32, and .73. A proaction design (Training A, Training B, and Test B) with 60 Ss per category was used. Training consisted of a list of 5 R_1 words followed by a list of 5 corresponding R_2 words. After 120 sec. of vowel cancellation, Ss were tested for recall of the R_2 words in the presence of the 5 stimulus words corresponding to the R_1 and R_2 words of training. Intrusions of R_1 words were found to be strongly related to the cultural probability of R_1. This finding was interpreted as the most solid empirical demonstration we have that language habits can interfere greatly in short-term retention.

Underwood and Postman (1960) have proposed that learned verbal material is subject to extraexperimental sources of interference. Extraexperimental sources refer to the individual's preexisting language habits which operate through proaction to interfere with the recall of more recently learned material. Indeed, Underwood (1957) suggested earlier that the major portion of forgetting is produced by proacting agents. To date, however, there has been no single, solid empirical demonstration that language habits do interfere greatly with the retention of laboratory learned words.

The source of interference to be investigated here was derived mainly from the work of Russell and Jenkins (1954), Bilodeau, Fox, and Blick (1963), and Blick (1963). The former published norms for response words given in discrete-free association to the 100 Kent-Rosanoff stimulus words. The stimulus word evoked a number of response words which were arranged in a hierarchial order of frequency. Bilodeau et al. made an analysis of these hierarchies and calculated the probability (p) relationships between the stimulus words and their response associates. The primary word was called R_1, the secondary word was called R_2, and responses from 3 to n were called R_{3-n} collectively. For example, CHAIR is an R_1 and its p of occurrence to the stimulus TABLE is .84. The p of R_2 (FOOD) is .04, and p of R_{3-n} is .12. Three categories of words were discovered in the norms. The probability of evoking R_1, R_2, or R_{3-n}, given the stimulus, is shown in parenthesis in Table 1. The categories have different patterns of probabilities. For example, Category A has low $p(R_1)$, low $p(R_2)$, and high $p(R_{3-n})$; Category B has medium $p(R_1)$, maximum $p(R_2)$, and medium $p(R_{3-n})$; and Category C has high $p(R_1)$, low $p(R_2)$, and low $p(R_{3-n})$. These patterns can be used to great advantage.

If S was trained with a list of R_1

[1] This article is an abridged version of a dissertation submitted to the Graduate School of Tulane University in partial fulfillment of the requirements for the PhD degree. The author wishes to acknowledge the assistance of E. A. Bilodeau and his project on memory supported in part by the National Science Foundation.

[2] Now at Randolph-Macon College.

words followed by a list of corresponding R_2 words, it would be possible for the R_1 words to interfere and intrude during a recall session for R_2 words. This would be especially true if stimulus words corresponding to the R_1 and R_2 words were presented during recall as a means of abetting the intrusions. Reminding during recall is a general method of controlling recall and has been discussed extensively elsewhere (Bilodeau, Levy, & Sulzer, 1963).

The amount of interference from primaries should vary in relation to the strength of $p(R_1)$. Accordingly, Category C should produce the greatest number of intrusions because its $p(R_1)$ is highest. Category B should produce a moderate number of intrusions, and Category A should produce the least number. The R_1 probabilities among the three categories were employed as a basis for predicting the amount of proactive interference produced by the cultural primary.

Method

Subjects and apparatus.—Ninety students from Louisiana State University and 90 male Naval enlisted reservists from the New Orleans area were unsystematically divided into three groups. A group contained 30 students and 30 reservists. Of the 90 college students, there were 57 males and 33 females, and the females had approximately equal representation from group to group.

The basic instrument was a 5-page booklet made from an opaque cover sheet and four sheets of ditto paper. The five sheets were arranged in the sequence: cover page, first training page, second training page, vowel cancellation page, and recall page.

Groups and procedure.—The three groups represented the three categories of words. A category contained 15 stimulus words which were unsystematically divided into three lists of five. The three lists within categories were represented by 20 *S*s each.

Following the proaction paradigm, the booklets contained five R_1 words and five R_2 words on the first and second training pages, respectively. The recall page contained the corresponding stimulus words and five empty lines for answers. Note that the R_1 and R_2 words of training represented the primary and secondary associates to the stimulus words written on the recall page.

Booklets and pencils were distributed to intact classes by assistant *E*s who served as proctors. To reduce the chances of *S* opening the booklet prematurely, the pages were secured by a paper clip. The lists were distributed in an unbiased fashion to reduce the possibility of copying. Page turning was paced by *E* who directed *S* when to turn and read the brief instructions printed at the top of the pages.

A total of 20 sec. per page was allowed for training, and on both pages the instructions read: "Study the words shown below." After training, *S*s were given 120 sec. to cancel vowels. The recall page read:

> A short while ago you studied some words on page —. The words below may help you remember those words. Write down the words that appeared on page —. You must fill in every line. If you cannot remember, write in the first word that the printed word makes you think of. Do not look back to the previous pages.

The recall phase lasted 120 sec.

The first and second training pages were labeled A and B, respectively, and the two blanks in the recall instructions were filled with Bs. The Bs were hand printed in red ink and were emphasized in *E*'s oral instructions so that *S* would not mistake the prescribed page.

Each word recalled was classified according to one of six sources of interference. Three sources consisted of Position in the Kent-Rosanoff Hierarchy (R_1, R_2, or R_{3-n}). The other three sources included Same-List Intrusion (a correct word given in the wrong place on the recall page), Blank (where *S* failed to respond), and Remainder (none of the previous five). Only R_1 and R_2 words written in the place corresponding to their position in the training list were scored as other-list intrusions (R_1) or as correct responses (R_2). When R_1 or R_2 appeared out of proper position, R_{3-n} took precedence in scoring. This rule provided a conservative estimate of both other-list and same-list intrusions.

Results

In Table 1 the 900 (180×5) pieces of recall data are classified by Category and Position in the Kent-

TABLE 1

OBTAINED AND NORMATIVE RESPONSE PROBABILITIES CLASSIFIED BY CATEGORY OF STIMULATION AND SOURCE OF RECALL

Category	Position in K-R Hierarchy		
	$p(R_1)$	$p(R_2)$	$p(R_{3-n})$
A	.21(.14)[a]	.51(.12)	.18(.74)
B	.40(.32)	.42(.24)	.07(.44)
C	.55(.73)	.29(.05)	.08(.22)

[a] Russell-Jenkins normative probabilities.

Rosanoff hierarchy. The value .21 entered under $p(R_1)$ for Category A means that 21% of the 300 recalled words consisted of intrusions from the other list, the value .51 under $p(R_2)$ means that 51% of the 300 words were correct, and so on.

The obtained probabilities show considerable agreement with the trends of the normative probabilities shown in parenthesis. The most important result occurred under $p(R_1)$ where the incidence of primaries increased as the normative $p(R_1)$ increased. The stimulation of recall has also raised the values of $p(R_2)$ over their original cultural incidence, though, Category B did not produce the maximum $p(R_2)$. Also, the data show a slight reversal under $p(R_{3-n})$ for Categories B and C, but a difference of .01 is minor.

The positive relationship between the mean number of intrusions of primaries (out of possible five) and the normative $p(R_1)$ is shown by the upper curve in Fig. 1. Analysis of variance for Category yielded a significant effect, F (2, 177) = 18.82, $p < .01$, and a comparison of the means by Duncan's range test indicated that the categories differed significantly from each other at the .05 level.[3] The two lower curves in Fig. 1 provide valuable control comparisons since they represent the same three categories taken from the Bilodeau, Fox, and Blick study which trained with *unrelated* R_2 words and no primaries. The lowest curve, which represents intrusions of R_2 words, reveals the lack of relationship between the number of other-list intrusions and $p(R_1)$ in marked contrast to the present findings. Not only do these data show no relationship but also the absolute values of intrusions are quite

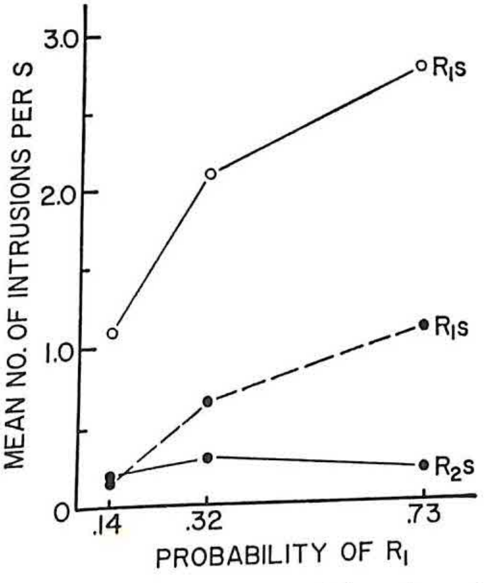


FIG. 1. Mean number of intrusions of primaries or secondaries as a function of the normative probability of R_1 for Categories A, B, and C.

TABLE 2

FREQUENCY DISTRIBUTION OF *S*s PRODUCING OTHER-LIST INTRUSIONS

Category	No. of Primaries Intruding						*S* Total
	0	1	2	3	4	5	
A	17	26	11	6	0	0	60
B	16	9	10	11	7	7	60
C	7	9	13	7	9	15	60

[3] The data for students and reservists were combined because all conclusions concerning intrusions as a function of $p(R_1)$ were the same for both groups.

low. This control comparison shows that other-list intrusions are low when the interpage relationship is low. The middle curve reflects the occurrence of primaries when *S*s are *not* trained with primaries and shows the level of interference produced solely by prelaboratory habits. The top curve, then, illustrates the combined effects of laboratory and prelaboratory habits on intrusions of R_1 at recall. Additional information regarding intrusions of R_1's is provided by a replication of the present study which omitted the stimulus words during recall. With the stimulus words absent, the occurrence of R_1's averages only 7% for the three categories.

A breakdown of the production of R_1's by *S*s is presented in Table 2. The strength of the effect of $p(R_1)$ in producing an R_1 is measured by the change in the distribution from category to category. As $p(R_1)$ increases, there is a corresponding increase in the number of *S*s giving more R_1's. For example, the percentage of *S*s giving four and five R_1's is 0, 23, and 40 for Categories A, B, and C, respectively. Note also that 15 of 60 *S*s in Category C have recalled all the items of the *incorrect* list and have offered them as items of the *correct* list.

Approximately 90% of the responses were found tabled in the Russell-Jenkins norms; a result which demonstrates the generalizability of the Minnesota norms. Only a small portion of the recall data (10% or less) was explained as a same-list intrusion, blank, and remainder. These sources of interference should be of far greater importance under other conditions.

Discussion

These data, in accordance with the previous study (Bilodeau, Fox, & Blick, 1963), demonstrate the power of culturally determined probability relationships to predict what other people will say in a retention test. The critical finding was that the number of intrusions of primaries increased as the cultural $p(R_1)$ increased. Deese (1959) was also able to predict verbal intrusions in immediate recall using cultural association strength as the independent variable. He used word lists composed of the first 12 terms (R_1 through R_{12} in the present notation) and discovered that the *stimulus* word intruded in recall. The stimulus word was predicted from its percent occurrence as a response associate to the R_1 to R_{12} words in a special side experiment using the R_1 to R_{12} words as stimuli. In a related study, Rothkopf and Coke (1961) were able to predict the recall of Kent-Rosanoff stimulus words from word-association norms. They concluded that frequency of recall of a stimulus word in free recall was a direct function of the number of cues provided by other stimulus words in the list. The cue number of a stimulus word was defined as the number of stimulus words in the list to which the word in question occurred as an association response.

It is desirable to place in perspective the magnitude of interference produced by the present technique. As much as 55% interference was produced and in a substantial number of *S*s intrusions of primaries were total (100%). In the study by Bilodeau, Fox, and Blick (1963) which trained with *unrelated* R_2 words and no primaries, the maximum production of primaries was smaller (22% or less). Many studies have used unrelated nonsense syllables and have produced about 1% interference and related nonsense syllables have produced about 5% interference. It was for the reason that interference is typically small and difficult to produce that the present study was undertaken. Since the present findings raise the obtained level of interference by an order of magnitude, an interference theory of forgetting becomes much more attractive than it was before.

The procedure of opposing primaries and secondaries during training and stimulating recall by means of a common

stimulus term could provide a substitute technique for testing the unit-sequence hypothesis proposed by Underwood and Postman (1960). According to this hypothesis, frequency of usage is the major factor responsible for producing interference. The assumption is that more frequently used words will have stronger associative connections than less frequently used words, and retention for high-frequency words will be poorer because the associative connections will recover with rest and interfere with recall. Experiments specifically designed to test the unit-sequence hypothesis have produced rather bland results (Postman, 1961, 1962; Underwood & Postman, 1960). It now seems that the probability relationships between a word and its response associates provide a better procedure for producing interference. Frequency of occurrence of words may not be as important a variable as the probability relationships between stimulus and response terms.

Finally, a question about the experimental design: Why was the first training page with its R_1's used at all? A trial of practice with R_1's was used in order to *increase* the p of an intrusion of R_1 since the purpose of the study was to maximize intrusions of R_1. If the purpose had been to study the sole effects of prelaboratory habits (Bilodeau, Fox, & Blick, 1963), no primary words would have been presented during training. As it stands, the design used is considered to enhance the probability of interference from preexperimental S-R habits. Requesting *S* to free associate to a stimulus during the test if he cannot remember the word studied may be ruled out of bounds by some students of memory. The method is unique, for it is a *blend* of retention and free-association procedures. It possesses none of the virtues of either when used alone, and is not intended to replace their separate use; but in combination tells us about the quantity and quality of misrecall under quantified conditions of modified free recall. Instructing *S* to use the recall stimulus leads to predictable patterns of cultural primary intrusions; exposing the primaries during training augments such intrusion patterns. These misrecalls may be indexes of the forgetting process itself or they may only be responses substituting after forgetting has occurred; the present design does not distinguish between these two alternatives. Most importantly, the design does show how particular misrecalls can be regulated in large quantities.

REFERENCES

Bilodeau, E. A., Fox, P. W., & Blick, K. A. Stimulated verbal recall and analysis of sources of recall. *J. verbal Learn. verbal Behav.*, 1963, 2, 422–428.

Bilodeau, E. A., Levy, C. M., & Sulzer, J. L. Long-term retention under conditions of artificially induced recall of related events. *Percept. mot. Skills*, 1963, **16**, 895–910.

Blick, K. A. *Dominant responses as a source of interference in verbal retention.* (Doctoral dissertation, Tulane University) Ann Arbor, Mich.: University Microfilms, 1963, No. 64–5915.

Deese, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22.

Postman, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, **61**, 97–110.

Postman, L. The effects of language habits on the acquisition and retention of verbal associations. *J. exp. Psychol.*, 1962, **64**, 7–19.

Rothkopf, E. Z., & Coke, E. U. The prediction of free recall from word association measures. *J. exp. Psychol.*, 1961, **62**, 433–438.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract No. N8onr-66216, Office of Naval Research.

Underwood, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

Underwood, B. J., & Postman, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

(Received January 16, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 251–256

SOME EFFECTS OF CONTOUR ON SIMULTANEOUS BRIGHTNESS CONTRAST [1]

PHYLLIS W. BERMAN AND H. W. LEIBOWITZ [2]

University of Wisconsin

Simultaneous brightness contrast was measured as a function of: (a) the orientation of a test object, shaped as a figure 8, on a half light, half black surround, (b) type and width of a contour separating the figure halves on the divided background. 48 adult Ss matched the brightness of the figure half on the dark background with that on the light surround. Subjective contrast was significantly greater: (a) when the figure 8 was presented with its rings on backgrounds of different brightness than when each ring lay on both backgrounds, (b) when figure halves were moved apart, each into its own surround, rather than when a dividing line separated the halves, (c) as width of the contour between halves was increased. The results are discussed in terms of the contribution of the border to subjective contrast obtained with complex stimulus configurations.

A gray stimulus viewed against a white background will appear darker than when viewed against a black background. This is an example of simultaneous brightness contrast, a phenomenon which can be interpreted in terms of neural inhibition at the retinal level following the implications of a number of recent electrophysiological experiments (Ratliff, Hartline, & Miller, 1963). Differences in excitation are exaggerated by interaction, the magnitude of the difference depending in a systematic manner on the luminance difference, size of background and test-fields, and the separation between the fields (Diamond, 1960, 1962; Heinemann, 1955).

There are, however, some aspects of the subjective contrast effect, such as its relation to contour, which do not fit readily into a theoretical approach based simply on neural inhibition. In a classical demonstration, Wundt showed that a gray paper placed over the border between a black and white area will tend to appear uniformly gray despite the expected effects of simultaneous contrast (Osgood, 1953). If however, a contour is added to the gray stimulus, so as to separate the portion viewed against the black from that viewed against the white background, the expected contrast effect is manifested. Helson (1943) discussed this phenomenon in terms of the "assimilation" of the contrast effect within a common border. Koffka (1935) theorized that shape acts as a "force of cohesion" which plays a role in the appearance of contrast. He suggested that if a gray object is used which has two clear subdivisions, degree of contrast will vary with its orientation on the contrast-inducing background. These phenomena suggest that subjective contrast depends, not only on luminance differences, but also on the presence of an identifiable contour or shape with respect to the stimulus.

The present experiments were designed to determine quantitatively the magnitude of the subjective con-

[1] Supported in part by Research Grant M-1090 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The authors wish to express their appreciation to R. B. Freeman, Jr. for assistance in the preparation of the manuscript.

[2] Now at Pennsylvania State University.

trast effect as a function of the width of the contour separating the two halves of a figure 8 test object, and as a function of the orientation of the test object on the black and white background. Such data should be of value in evaluating the importance of contour as a contributing variable to the phenomenon of simultaneous brightness contrast.

Apparatus

The stimulus consisted of a figure 8 cut from neutral density film (D = 0.7) and mounted flush against a piece of vertically positioned milk glass. Illuminance was provided by a 300-w., 110-v. ac projector system focused on the opposite side of the milk glass. The right half of the stimulus (see Fig. 1) was positioned within a clear rectangle and appeared less luminous than its background. The left half of the figure was surrounded by opaque black paper and appeared more luminous than its background. Except for the figure 8 and the light background surrounding the right half of the figure, there were no other stimuli in the visual field. The remaining portions of the milk glass were covered with opaque black paper.

The *S* was able to match the two halves of the figure 8 by means of a pair of polaroid filters. The polarizer, made from sheet polaroid, was placed over the figure half lying on the black background. The analyzer, a similar piece of polaroid mounted in glass, was placed in the eyepiece through which *S* observed monocularly with the natural pupil. The *S* rotated the analyzer until the two figure halves appeared equal. The matched luminance value of the left half of the figure could be calculated from the angular position of the analyzer and the calibration data of the polaroid system.

The figure 8 stimulus was symmetrical horizontally and vertically with an outside diameter of each ring of $2\frac{1}{4}$ in. The rings were $\frac{3}{8}$ in. width. The viewing distance was 34.5 in. at which the maximum dimensions of the figure 8 were 3.72 × 7.44° of arc. The bright background against which the right half of the figure was viewed subtended a visual angle of 13.7 × 5.8° of arc.

The magnitude of the contour was varied by presenting the stimulus with no inserts between the two halves of the figure, and by placing vertical wires or strips of the following widths symmetrically over the center of the figure: .005, .01, .02, .04, .08, .16, .32, and .64 in. These dividers corresponded to visual angles of 0.5, 1.0, 2.0, 4.0, 8.0, 15.9, 31.9, and 63.7′ of arc. The luminance of the right half of the test figure was constant at log 2.0 mL. and the background luminance was constant at log 2.7 mL.

Experiment I

Method

Two groups of 12 *S*s each, volunteers with normal vision from the elementary psychology classes at the university, first made four practice matches using a square test object and a dividing strip subtending 63.7′ of arc. Group A was then presented the figure in the vertical position, and Group B in the horizontal position (see Fig. 1). The nine experimental conditions for each group, i.e., eight dividers and no divider, were counterbalanced by a 9 × 9 Latin-square design. Subjects 10, 11, and 12 were presented with the same sequence as *S*s 1, 2, and 3 since ordinal

TABLE 1

Average Log Matched Luminance in Millilamberts of Groups A, B, C, and D as a Function of Width Divider or Separation between Figure Halves

Group	Width Divider in Minutes of Arc								
	0	0.5	1.0	2.0	4.0	8.0	15.9	31.9	63.7
A	1.13	0.97	0.95	0.78	0.80	0.65	0.31	0.13	−0.15
B	0.08	−0.20	−0.15	−0.10	−0.22	−0.24	−0.39	−0.51	−0.71
C with separation	0.92						−0.15	−0.25	−0.43
C with divider	0.92						0.49	0.12	−0.08
D with separation	0.12						−0.45	−0.62	−0.64
D with divider	0.12						−0.43	−0.51	−0.80

Note.—Groups A and B are from Exp. I. Groups C and D are from Exp. II.

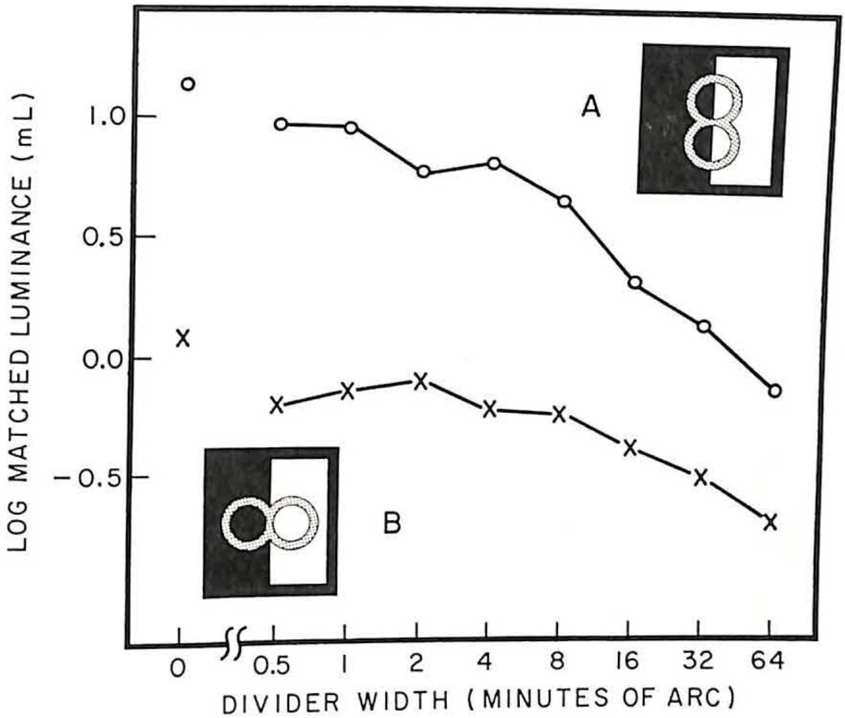


FIG. 1. Log matched luminance of the left half of a figure 8 test object viewed against a black background matched with the right half viewed against a light background. (The background luminance was constant at log 2.7 mL. The luminance of the right half of the figure was constant at log 2.0 mL. The data in the upper curve, indicated by circles, were obtained with vertical orientation of the configuration [Insert A; no divider present]. The data for the lower curve, indicated by crosses, were obtained with a horizontal orientation [Insert B].)

position, as indicated by analysis of variance, was not a significant variable. The *S* made four matches at each of the experimental conditions.

Each *S* was given a period of preliminary dark adaptation after which instructions were given regarding the nature of the task and the method of rotating the polaroid in the eyepiece so as to produce a match between the two halves of the test object. The *S*s were instructed not to try to match any one section of the figure with any other section, but to make the best overall match possible.

Results

The data are presented in Table 1 and plotted in Fig. 2 as the log matched luminance as a function of width of the divider for the two orientations used. Table 2 is a summary of an analysis of the variance associated with these data. All terms tested in this analysis are significant at the .001 level.

Even in the absence of any divider between the two halves of the test field, indicated by the zero position on the abscissa axis of Fig. 2, considerable contrast existed. Whereas

TABLE 2

ANALYSIS OF VARIANCE OF MATCHING SCORES OF GROUPS A AND B USING NINE WIDTH DIVIDERS

Source	*df*	*MS*	*F*
Between *S*s			
Figure position (FP)	1	427,645.0	19.14***
Error (b)	22	22,339.1	37.34***
Within *S*s			
Width divider (WD)	8	25,994.8	43.46***
WD × FP	8	2,827.4	4.73***
Error (w)	176	598.2	
Total	215		

*** $p < .001$.

the luminance of the portion of the figure viewed against the white background was log 2.0 mL., it was matched by the portion viewed against the black background at an average value of log 1.1 mL. for the vertical, and log 0.1 mL. for the horizontal orientation.

The effect of introducing the divider is considerable, amounting to factors of approximately 16 to 1 for the vertical, and 5 to 1 for the horizontal orientation. As dividers of greater width were introduced the amount of contrast was systematically increased, indicated by the increasingly lower luminance values of the left half of the figure required to produce a match between the two halves of the test figure. This effect was exhibited with the test object in both orientations, but there was significantly more contrast for all divider widths with the test figure oriented horizontally than with the vertical orientation.

Increasing the width of the divider had a greater effect with the test figure in a vertical position. This can be seen in Fig. 1 and is reflected by the significant interaction between Width Divider and Figure Position in the analysis of variance presented in Table 2. An extended Alexander trend test was used to determine whether the increase in contrast was significantly more linear with the test figure in the vertical position than with it in the horizontal position (Grant, 1956). The interaction term was broken into the component terms, Between Group Slopes, Quadratics, Cubics, and Deviations from the Expected. Of these, only Between Group Slopes is significant, $F\ (1, 22) = 8.86$, $p < .01$.

TABLE 3

ANALYSIS OF VARIANCE OF MATCHING SCORES OF GROUPS C AND D WITH .160-, .320-, AND .640-IN. DIVIDERS AND SEPARATION BETWEEN HALVES

Source	*df*	*MS*	*F*
Figure position (FP)	1	99,487.6	15.93***
Ss/FP	22	6,244.6	33.97***
Width between halves (W)	2	15,071.7	32.76***
Between divider and separation (D-S)	1	17,844.5	16.35***
FP × D-S	1	19,716.9	18.07***
W × FP	2	745.9	1.62
W × D-S	2	1,683.5	9.16***
W × FP × D-S	2	5,968.4	32.47***
Ss/FP × W	44	460.1	2.50**
Ss/FP × D-S	22	1,091.1	5.94***
Ss/FP × W × D-S	44	183.8	
Total	143		

Note.—*Ss*/FP is the error term for FP. *Ss*/FP × W is the error term for W and for W × FP. *Ss*/FP × D-S is the error term for D-S and for FP × D-S. *Ss*/FP × W × D-S is the error term for all other terms.

** $p < .01$.
*** $p < .001$.

EXPERIMENT II

Method

In Exp. I the dividing strip covered part of the test object so that an increase in the width of the divider decreased the effective area of the test object as well as the background. Since area is an important variable in contrast (Diamond, 1960, 1962), a control experiment was conducted in which the two halves of the test object were physically separated so that, while the width of the "contour" between them increased as in Exp. I, their areas, as well as the area of the light background, remained constant. Two additional groups of 12 *S*s each were tested under conditions of no divider, and with dividers as well as separations of .16, .32, and .64 in. Group C was presented with the test object in the vertical position, and Group D with it in the horizontal position.

The order of presentation of the seven treatment conditions was determined by a 7 × 7 Latin-square design with the same design being repeated for both groups. The results of a *t* test between the most divergent scores indicated that ordinal position was not significant, so that the five additional *S*s in each group were presented with the same order of treatment conditions as *S*s 1, 2, 3, 4, and 5.

Results

The data are presented in Table 1 and Table 3 presents an analysis of the variance of these data. Since only

one measurement per S represents the zero level of both the Divider and Separation conditions, only the scores for the .16, .32, and .64 widths were included in this analysis.

As in Exp. I, both Width and Figure Position are significant sources of variability.

The overall effect of separation of the halves of the test figure was to produce significantly more contrast than was perceived when a dividing strip was used to create a contour between the figure halves. The significance of the interaction term, Figure Position × Divider Separation, reflects the fact that this effect differed for the two orientations of the test figure used. Separation produced more contrast when the test figure was in the vertical position. For the horizontal orientation, either procedure produced similar results.

In the analysis summarized in Table 3 Width × Figure Position is insignificant but the second-order interaction, Width × Figure Position × Divider Separation, is significant. Analyses of variance were computed separately for the Separation condition and for the Divider condition and data for the zero level of each was included. Width × Figure Position was significant only for the Separation condition, $F\,(3, 66) = 4.60$, $p < .01$. This finding differs from the result in Exp. I, that Width Divider × Figure Position is significant, but with nine instead of three widths dividing strips used.

Width is a significant variable in both the Separation condition, $F\,(3, 66) = 62.83$, $p < .001$, and the Divider condition, $F\,(3, 66) = 36.39$, $p < .001$. However, the increase in contrast perceived with increasing widths between figure halves differed in the Separation and Divider conditions. This is indicated by the significant interaction, Width × Divider Separation, in Table 3.

Discussion

The results of the present experiments provide quantitative confirmation of the phenomenal impressions derived from early demonstrations. The contribution of border to the subjective contrast effect is clearly demonstrated by the marked change in the magnitude of contrast resulting from the introduction of thin dividing strips between the halves of the test object. Of particular interest is the influence of the smallest divider used in Exp. I, which subtended only 30″ of arc. This value corresponds to the angle subtended by individual cones in the center of the human fovea, and is approximately the same order of magnitude as the threshold for grating acuity (Leibowitz, 1953). This fine line is seen here to be adequate to increase contrast for both orientations of the figure. It is highly unlikely that the effect of a line of this width, which represents approximately 0.2% of the total width of the test object in the vertical orientation and 0.1% of the total width in the horizontal orientation, can be ascribed to changes in luminance or area relationships. The finding that there is not less but more difference in brightness between figure halves in the Separation condition indicates that the findings cannot be attributed to changes in area resulting from introduction of dividing strips. A more plausible explanation would be in terms of an "equalizing mechanism" such as that described by Fry (1948; Fry & Bartley, 1935), which is assumed to be responsible for the assimilation of differential contrast effects throughout the whole figure or, when the divider is introduced, within the figure halves.

When there is no divider introduced, the shape of the figure itself may have an influence similar to the effect of such a contour. The extent to which this is manifested depends on the position of the test object on the divided background. Because the figure 8 test object is of such a shape that two component

parts are identifiable, the position of these two parts on the divided background is important. More contrast is observed between the figure halves viewed against different backgrounds than when the rings of the figure each lie on both the black and white backgrounds, as in the vertical orientation.

Of interest is the observation that considerable subjective contrast exists in the absence of any dividing line between the two halves of the figure on backgrounds of different brightness. The effect is probably not noticeable in the more familiar demonstrations of this phenomenon using paper stimuli. With such stimuli contrast appears to be reduced by the presence of the visible microstructure of the paper surface, and it may be increased by observing test objects through a diffuser such as tissue paper (Woodworth & Schlosberg, 1955). In the present stimulus arrangement transmitted light was utilized and no microstructure was visible. In addition the luminance differences among the test object and the light and dark backgrounds with transmitted light are far in excess of that available with paper stimuli. Thus, the present arrangement results in far more favorable conditions for contrast than ordinarily exist. The familiar observation that a divider is necessary for contrast to be manifested seems to be a special case resulting from the use of paper stimuli.

In the present experiments the role of the shape of the test object and of fine dividing lines has been demonstrated. It follows that simultaneous contrast is not a simple function of luminance and spatial variables. This does not imply that these variables are not fundamental to contrast, but rather that additional concepts are needed to fully explain subjective contrast with more complex stimulus configurations.

REFERENCES

DIAMOND, A. L. A theory of depression and enhancement in the brightness response. *Psychol. Rev.*, 1960, **67,** 168–198.

DIAMOND, A. L. Simultaneous contrast as a function of test-field area. *J. exp. Psychol.*, 1962, **64,** 336–345.

FRY, G. A. Mechanisms subserving simultaneous brightness contrast. *Amer. J. Optom.*, 1948, **45,** 1–17.

FRY, G. A., & BARTLEY, S. H. The effect of one border in the visual field upon the threshold of another. *Amer. J. Physiol.*, 1935, **112,** 414–421.

GRANT, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53,** 141–154.

HEINEMANN, E. G. Simultaneous brightness induction as a function of inducing- and test-field luminances. *J. exp. Psychol.*, 1955, **50,** 89–96.

HELSON, H. Some factors and implications of color constancy. *J. Opt. Soc. Amer.*, 1943, **33,** 555–567.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

LEIBOWITZ, H. Some observations and theory on the variation of visual acuity with the orientation of the test object. *J. Opt. Soc. Amer.*, 1953, **43,** 902–905.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

RATLIFF, F., HARTLINE, H. K., & MILLER, W. H. Spatial and temporal aspects of retinal inhibitory interaction. *J. Opt. Soc. Amer.*, 1963, **53,** 110–120.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1955.

(Received January 17, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 257–262

PREEXPOSURE TO VISUALLY PRESENTED FORMS AND NONDIFFERENTIAL REINFORCEMENT IN PERCEPTUAL LEARNING [1]

LARRY C. KERPELMAN [2]

University of Rochester

An experimental clarification of previous preexposure experiments which seemed to support the differentiation viewpoint of perceptual learning was undertaken. 79 albino rats were reared from birth in a visually sparse environment. Circles and triangles hung in the cages of the experimental Ss. Nondifferential food reinforcement of these cage stimuli was varied, and the resultant effect on subsequently learning to discriminate between them was examined. Discrimination performance of the main experimental groups confirmed the hypothesis that nondifferential reinforcement played a role in previous preexposure studies of perceptual learning, thus supporting the enrichment approach. Results for supplementary groups necessitated expansion of that approach to include possible reinforcing effects of novelty.

Several studies in which rats were exposed to stimuli (usually circles and triangles) from birth and later learned to discriminate between them better than nonexposed *S*s (Gibson & Walk, 1956; Gibson, Walk, Pick, & Tighe, 1958; Gibson, Walk, & Tighe, 1959; Walk, Gibson, Pick, & Tighe, 1958, 1959) have been used to support the "differentiation" view of perceptual learning. That view (Gibson & Gibson, 1955) holds that experience with stimuli that are to be discriminated is all that is necessary in certain situations for perceptual learning to occur. It is not necessary to associate drive reduction, memory, or other factors with the stimuli. The Gibsons have named the opposing viewpoint, which subsumes association, reinforcement, and drive-reduction theories, "enrichment." The differentiation interpretation of the experiments cited above (which will hereafter be referred to generically as the Cornell studies) is exemplified by the statement: "The theories which best seem to fit the results all emphasize experience per se rather than *reinforced* experience [Walk et al., 1958, p. 487]."

An alternative explanation of those results is nevertheless plausible. In the Cornell studies, experimental *S*s received food, water, and other reinforcement in the presence of the cage forms to which they were preexposed. Hence the stimuli were nondifferentially reinforced. Spence (1936) postulated that nondifferential reinforcement of, for example, two stimuli raises the excitatory tendency of both. On introducing differential reinforcement (as in the discrimination task), enhancement of learning would be expected in *S*s receiving prior nondifferential reinforcement (the preexposed Cornell experimental *S*s) as compared to those receiving little or

[1] This article is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Rochester. The author is greatly indebted to Norman Harway for his patient guidance and encouragement: the valuable assistance and suggestions of Robert Ader, Robert Boynton, and G. R. Wendt are also gratefully acknowledged. A condensed version of this paper was presented at the August 1963 Psychonomic Society meetings in Bryn Mawr, Pennsylvania.

[2] Now at the Institute for Juvenile Research, Chicago.

no prior reinforcement—differential or nondifferential—contiguous with the stimuli (the nonexposed Cornell control *S*s). By this formulation the preexposed *S*s in the Cornell experiments might be expected to learn the subsequent discrimination faster than the controls because, at the beginning of discrimination training, the stimuli for the former group were at a higher level of excitatory tendency than they were for the latter group. Because the Cornell studies have important implications for perceptual theories, the present study was undertaken. The effect of nondifferential reinforcement in a preexposure situation was examined by varying the presumed strongest reinforcement, food, within conditions replicating, in all essentials, those of the Cornell studies.

Method

Subjects

The *S*s were 79 Charles River CD albino rats, 40 in Replication 1 (R1) and 39 in Replication 2 (R2). These were selected from a pool of 105 rats born to 10 pregnant females from the Charles River Breeding Laboratories. The 55 rats providing the R1 *S* pool were all born within a span of 5 hr.; the 50 rats providing the R2 *S* pool were all born 47 days later within a span of 6 hr.

Rearing

The *S*s were born and reared in cages made of $\frac{1}{2}$-in. wire mesh, the cages placed in white compartments dimensionally identical to those of Walk et al. (1958). The backs of the compartments were flush with the white wall of the room, and the front was covered by a white oilcloth curtain hanging 8 in. from the edge of the compartments. Water was available ad lib. Thus the visual environment of each *S* during the experiment consisted of the compartment, the cage, the water bottle, the cage mates, the food tray for certain periods each day, and, for the experimental *S*s, the cage stimuli for certain periods each day. Each of the R1 cages was lit from 7:00 A.M. to 6:00 P.M. daily by a shielded 10-w. bulb on the compartment ceiling; R2 cages had no such arrangement but received normal room- and daylight during the same period.

Approximately 1 day after birth the five litters of each replication were cross-fostered, each mother receiving an equal number of pups. Each group thus formed was assigned to one of the four experimental conditions or to the control condition. The *S*s nursed ad lib. until age 21 days when they were weaned and sexed. Each sex within each condition was placed in a separate cage, three to seven *S*s per cage. A 12-hr. feeding schedule was begun, and, for the experimental groups, stimulus-placement procedures (discussed below) were initiated. Except for cross-fostering and weaning, *S*s were not handled during rearing.

After weaning, *S*s were fed twice daily with a mash of powdered laboratory meal mixed with water. The food was placed in white food trays that were hung from the cage wall for feeding and then removed after 1 hr. The first feeding was at approximately 8:00 A.M., the second, at approximately 7:00 P.M.

Four stimuli, two circles and two triangles, hung in the experimental *S*s' cages, one against each wall of each cage, for a period of time daily (discussed below). They were made of flat $\frac{1}{32}$-in. brass painted flat black. The circles were 3 in. in diameter, the triangles, $3\frac{1}{2}$ in. per side. Soldered to the top of each form was a length of wire by which the stimuli were hung in the cage. They were rotated at each feeding so that a circle and a triangle hung against each wall each day, and, specifically, a circle or a triangle was alternately at the wall where *S*s received food each day.

Experimental Conditions

Stimulus placement procedures were introduced at age 21 days because the ad lib. nursing allowed up to that time precluded control of feeding and stimulus perception—especially when *S*s' eyes opened at age 16 days—as required by the experimental conditions. These procedures continued throughout the experiment. For Group E1 (22-Hr. Reduced Reinforcement), the four cage stimuli were present 22 hr. daily, being removed for 2 hr. daily at each 1-hr. feeding period. This group provided the needed conditions to discern the factors involved in the improved discrimination performance of preexposed groups in the Cornell studies. For Group E2 (22-Hr. Nondifferential Reinforcement), the cage forms were present 22 hr. daily, being removed twice daily at separate 1-hr. nonfeeding times to equate total amount of preexposure with that of Group E1. Since the forms were present during feeding, this group was comparable to the typical Cornell experimental group.

The *S*s of Group E1 might have been open to reinforcement factors other than the manipulated one, food, such as social reinforcement, secondary reinforcement, and primary reinforcement of the thirst drive. Because these might have confounded the food reinforcement variable, supplementary groups were instituted which were expected not to be as prone to such influences. Group E1′ (2-Hr. Reduced Reinforcement) was exposed to the forms for 2 hr. daily at separate 1-hr. nonfeeding times. Group E2′ (2-Hr. Nondifferential Reinforcement) was exposed to the cage forms only for the 2 hr. daily that feeding occurred.

For Group C (Control), no cage forms were present during the experiment. This group was comparable to the Cornell studies' control groups.

In Groups E2 and E2′, where the stimuli were present during feeding, a stimulus was placed directly over the food tray. The mash was of such consistency that it could not be carried from the tray and thus had to be eaten in close proximity to the stimulus.

In an attempt to control a possible confounding factor, viz. differences in activity level at the time of stimulus change, stimuli were removed from E2 cages and placed in E1′ cages the hour before feeding as often as possible (47% of the time in R1 and 30% in R2). This was done to equate activity level of these groups at the time of stimulus change with that of Groups E1 and E2′, where stimulus change occurred at feeding, a period of high activity (Reid & Finger, 1955). It was expected that all four experimental groups would thereby "notice" the stimulus change equally because of approximately equal activity level. When the stimuli were removed from E2 cages and placed in E1′ cages the hour before feeding, an interval of approximately 2 min. was interposed between removal of the stimuli from E1′ cages and their replacement in E2 cages, and the placing of the food trays in the cages. This was done to dispel the signaling effect that stimulus change may have had: the limited memory capacity of the albino rat (Honzik, 1931) made it seem unlikely that a relationship between stimulus replacement (E2) or removal (E1′) and appearance of food would be made.

A single 1-hr. daily feeding was begun at age 86 days in preparation for the circle-triangle discrimination learning task at age 90 days. This 24-hr. schedule was continued until the end of the experiment, and appropriate adjustments were made in stimulus placement procedures during this period.

Discrimination Training

The apparatus was a modified Grice (1948) discrimination apparatus as described elsewhere (Gibson & Walk, 1956). Pretraining and discrimination training followed the procedures described in Gibson and Walk (1956). The *S* could obtain a pinch of wet mash from one of two stimulus panels—a black circle, or a black triangle, on a white background—by pushing open the correct stimulus door. The incorrect stimulus panel was baited but locked. The only differences between the Gibson and Walk (1956) procedure and the present one was that a 60-sec. intertrial interval was not adhered to, and *S*s were run for 40 days if they had not reached criterion, rather than 15 days. (See Kerpelman, 1963 for further procedural details.)

RESULTS

The 79 *S*s chosen from the pool of 105 animals were selected before discrimination training to give as equal a distribution as possible within conditions, replications, and sexes. The resultant *N* was 16 in each group, except for E1 in which there were 15 *S*s due to inequality of sexes. For the statistical analysis, however, the missing *S* in Group E1 was "replaced" by adding the appropriate mean subgroup score. Means and *SD*s of the four performance measures for all subgroups[3] are given in Table 1. Since the subsequent analyses of variance did not reveal a significant Conditions × Replications interaction, the scores given and the analyses of differences between them are based on both replications combined.

A summary of the 5 × 2 × 2 anal-

[3] A mean and *SD* table for the subgroups of each replication (Table A), four analysis of variance summary tables (Tables B through E), and a *t*-test summary table (Table F) have been deposited with the American Documentation Institute. Order Document No. 8182 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1

MEANS AND *SD*s OF THE TREATMENT CONDITIONS FOR R1 AND R2 COMBINED

Groups	Trials to Criterion		Initial Errors		Repetitive Errors		Total Errors	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
E1	267.13	82.77	108.25	40.21	136.81	82.02	245.06	119.94
E2	188.63	90.12	71.25	38.63	82.50	55.40	153.75	83.98
C	249.69	63.37	97.69	30.84	108.88	49.60	206.56	69.03
E1′	205.25	70.75	75.25	31.80	80.44	53.42	155.69	83.06
E2′	189.31	42.45	70.19	19.43	78.50	30.91	148.69	46.75

yses of variance (Conditions × Positive Stimuli × Replications) performed on each measure is given in Table 2. The level of statistical significance used throughout was $p < .05$. In interpreting the results it should be noted that the four measures are related: significant results across all four are not, therefore, independent failures to support the null hypothesis.

A significant Conditions main effect is readily apparent. A Replications main effect is strongly suggested, manifested by all groups in R1 performing better than their counterparts in R2, with the exception of Group E2′. A Stimulus main effect is slightly suggested, shown by a tendency for *S*s for which circle was positive to make more errors, except, again, for *S*s in Group E2′. Conditions × Sexes × Replications analyses of variance, with cell frequencies adjusted for equal *N*, were done on all measures. The only significant findings were a Conditions main effect for all measures and a Sexes × Replications interaction for two measures, the interaction most likely reflecting confounding of sex with time of day of running.

In light of the consistently significant Conditions main effect, *t* tests of differences between condition means relevant to the research hypotheses were performed, the estimated error variance being based on the error mean square of the Conditions

TABLE 2

ANALYSES OF VARIANCE OF FOUR PERFORMANCE MEASURES

Source	*df*	Trials to Criterion		Initial Errors		Repetitive Errors		Total Errors	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Conditions (C)	4	20983	3.92**	4816	4.41**	10208	3.09*	28721	4.07**
Stimuli (S)	1	16245	3.04	3432	3.15	19035	5.76*	38632	5.47*
Replications (R)	1	8736	1.63	5746	5.27*	15569	4.71*	40231	5.70*
C × S	4	9396	1.76	1647	1.51	4237	1.28	10432	1.48
C × R	4	4476	<1	1106	1.01	2395	<1	5468	<1
S × R	1	13	<1	106	<1	719	<1	1378	<1
C × S × R	4	4163	<1	633	<1	60	<1	1036	<1
Within	59[a]	5351		1091		3305		7061	
Total	78[a]								

[a] 1 *df* subtracted because of "replaced" E1 *S*.
* $p < .05$.
** $p < .01$.

TABLE 3

RESULTS OF t TESTS RELEVANT TO THE RESEARCH HYPOTHESES IN TERMS OF STATISTICAL SIGNIFICANCE

Measure	22-Hr. Groups		2-Hr. Groups	
	Differentiation Prediction $E2 = E1 > C$	Enrichment Prediction $E2 > E1 \geq C$[a]	Differentiation Prediction $E2' = E1' = C$	Enrichment Prediction $E2' > E1' = C$
Trials to criterion	$E2 > E1 = C$		$E2' = E1' = C, E2' > C$	
Initial errors	$E2 > E1 = C$		$E2' = E1' = C, E2' > C$	
Repetitive errors	$E2 = C = E1, E2 > E1$[b]		$E2' = E1' = C$	
Total errors	$E2 = C = E1, E2 > E1$		$E2' = E1' = C$	

[a] Due to the possible operation of uncontrollable reinforcement contingencies in the presence of the stimuli in the absence of food, exact a priori predictions of the E1-C difference could not be made within the enrichment framework.

[b] Sign > indicates significantly better performance. Sign = indicates performance levels did not differ significantly, although they may have differed nonsignificantly. For example, $E2 = C = E1$ means E2 did not differ significantly from C, nor C from E1, yet E2 still may have differed significantly from E1.

× Stimuli × Replications analyses. The results, in terms of statistical significance, are summarized in Table 3.

DISCUSSION

22-Hr. Groups

The 22-hr. groups were comparable to the typical Cornell experimental group. The only difference in experimental conditions between Groups E1 and E2 was that, for the latter, feeding occurred in the presence of the cage forms; whereas for the former, the stimuli were absent during feeding. Yet E2 performed on the discrimination task significantly better than E1 and, on two measures, than C. The results suggest that the facilitating effect of preexposure to visually presented forms found in the Cornell studies derived predominantly from the nondifferential reinforcement *S*s received in the presence of those forms. That the results of those studies are explicable in terms of reinforcement theory tends to weaken the empirical support for the differentiation position. Furthermore, Spence's (1936) specific hypothesis concerning the effects of nondifferential reinforcement received support from these findings.

2-Hr. Groups

While the E2′-C differences on two measures were in line with enrichment predictions, if food reinforcement were the only factor accounting for learning enhancement on the discrimination task, then Group E2′ would have been expected to perform better than Group E1′. The values of the performance means were in this direction, but the differences were not significant. These results are at least as much in line with differentiation as with enrichment predictions. If, in the differentiation framework, 2 hr. per day were not enough time for one group to differentiate the stimulus elements, then in neither group would facilitation of learning be expected. On the other hand, differentiation theory cannot explain why E2′ performed better than C on two measures while E1′ did not, since both 2-hr. groups had the same amount of experience with the stimuli.

Novelty and Stimulus Change

The apparent inability of either position to account for the results of the 2-hr. groups can be resolved by introducing an additional explanatory concept, viz. the (nondifferentially) reinforcing effects of novelty and stimulus change. This concept postulates, consistent with recent work (Berlyne, 1960; Butler & Harlow, 1954; Dember, 1960; Forgus, 1958), that not only can food reinforcement account for the raising of excitatory tendency, but also reinforcement due to novelty and perceptual change can act

similarly. Applying this concept, the cage forms in Group E1′, appearing as they did for only two 1-hr. periods daily, were novel stimuli. It may be postulated that *S*s were attracted to them, in that their appearance satisfied the exploratory, change-seeking drive (Dember, 1960). As a result, the performance level of Group E1′ was brought up to a level wherein significant differences between the 2-hr. groups were eradicated.

If perceptual change is postulated to have reinforcing properties, the question of why E1′ did not perform significantly better than C (as did E2′ on two measures) must be approached. It can only be assumed that the effects of this kind of reinforcement are quantitatively not as great as the corresponding mechanisms for hunger. The "novelty reinforcement" group (E1′), while performing well enough to erase significant differences between it and E2′, did not perform well enough to differ significantly from C. This same assumption about the relative strength of novelty and food reinforcement holds for the failure to find significant differences between Groups E2 and E1′ as well (two-tailed t tests between these groups on all measures were not significant).

Future research might test the validity of the novelty hypothesis by varying the amount of daily form preexposure, divorced from food reinforcement. It would be hypothesized that the longer a stimulus is viewed, the less it will continue to reinforce the perceptual change drive. Two extremes of such a situation are the groups in the present study in which the stimuli were present in the absence of food. Group E1′, in which the stimuli were present 2 hr. daily, performed better than Group E1, in which the stimuli were present 22 hr. Two-tailed t tests on all four measures were significant ($p < .025$), thus supporting the novelty hypothesis.

REFERENCES

Berlyne, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

Butler, R. A., & Harlow, H. F. Persistence of visual exploration in monkeys. *J. comp. physiol. Psychol.*, 1954, **47**, 258–263.

Dember, W. N. *The psychology of perception.* New York: Holt, 1960.

Forgus, R. H. The effects of different kinds of form pre-exposure on form discrimination learning. *J. comp. physiol. Psychol.*, 1958, **51**, 75–78.

Gibson, E. J., & Walk, R. D. The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *J. comp. physiol. Psychol.*, 1956, **49**, 239–242.

Gibson, E. J., Walk, R. D., Pick, H. L., Jr., & Tighe, T. J. The effect of prolonged exposure to visual patterns on learning to discriminate similar and different patterns. *J. comp. physiol. Psychol.*, 1958, **51**, 584–587.

Gibson, E. J., Walk, R. D., & Tighe, T. J. Enhancement and deprivation of visual stimulation during rearing as factors in visual discrimination learning. *J. comp. physiol. Psychol.*, 1959, **52**, 74–81.

Gibson, J. J., & Gibson, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.

Grice, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, **38**, 633–642.

Honzik, C. H. Delayed reaction in rats. *U. Calif. Publ. Psychol.*, 1931, **4**, 307–318.

Kerpelman, L. C. Pre-exposure to visually presented forms and nondifferential reinforcement in perceptual learning. Unpublished doctoral dissertation, University of Rochester, 1963.

Reid, L. S., & Finger, F. W. The rat's adjustment to 23-hour food deprivation cycles. *J. comp. physiol. Psychol.*, 1955, **48**, 110–113.

Spence, K. L. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

Walk, R. D., Gibson, E. J., Pick, H. L., Jr., & Tighe, T. J. Further experiments on prolonged exposure to visual forms: The effect of single stimuli and prior reinforcement. *J. comp. physiol. Psychol.*, 1958, **51**, 483–487.

Walk, R. D., Gibson, E. J., Pick, H. L., Jr., & Tighe, T. J. The effectiveness of prolonged exposure to cutouts vs. painted patterns for facilitation of discrimination. *J. comp. physiol. Psychol.*, 1959, **52**, 519–521.

(Received January 17, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 263–268

DIFFERENTIAL VISUAL FEEDBACK OF COMPONENT MOTIONS

JOHN D. GOULD[1]

IBM Watson Research Center, Yorktown Heights, New York

New closed-circuit television techniques allowed the joint action of S's hand, control-instrument, and operational effects to be visually fed back singly or in combination. 8 visual feedback conditions and 2 levels of task precision were used. The results showed that the effects of visual feedback were determined by the component motions fed back, with the operational effects being the most important followed by control-instrument and hand-arm movements. A significant interaction between visual feedback and precision of movement occurred, where feedback of the operational component became more important as more overall precision of movement was demanded.

One way of considering behavior is in terms of sensory feedback, and a proper experiment within this framework is to assess performance while controlling such feedback. This research was concerned with the relationships of vision and aided motion and studied the effects of visual feedback of the different component motions involved in a complex motor control task. In all aided motion there are three main movement components: differentiated body movements such as those of the fingers, arm, hand, foot, or torso; movement of the tool, instrument or control; and movement of the objects or material being controlled or changed. The sensory feedback resulting in all modalities from these three nonindependent component movements may be designated as *reactive*, *instrumental*, and *operational* feedback, respectively (Gould, 1964; Smith, 1962).

Previous studies have shown that modifications in one or more sense modality of any one of the three main feedback mechanisms may indeed affect behavior. In the tracking literature, for example, evaluation of kinesthetic cues from various types of controls or instrumental effects while keeping other feedback variables constant showed significant differences (Chernikoff, Birmingham, & Taylor, 1955; Howland & Noble, 1953). Likewise, the differences between pursuit and compensatory tracking (Briggs, Fitts, & Bahrick, 1957) were due to visual operational feedback effects, while directionality of movement relative to the operator for a given control changed the effects of reactive feedback (Grether, 1947).

The present study was concerned with modification of the *visual* feedback of the three component motions. To accomplish this, a closed-circuit television system was designed which allowed real-time visual feedback of any one of the three component motions in a tool-using task, singly or in various combinations, where *S*s used a pliers to move a set of pins into prescribed holes. This technique allowed vision, for example, of the pliers without the pins or hand of the *S* being seen.

Two hypotheses were tested. The relative effects of vision are dependent upon the precision of the overall motion pattern involved; i.e., the more precise the overall motion pat-

[1] The critical comments of H. William Morrison in reading this paper have been appreciated.

tern the more behavior should be affected by reducing visual feedback. This hypothesis was tested by varying the size of "target" holes to which the pins had to be moved. Secondly, for a given overall motion pattern lack of vision of the most complex of the three motion components should affect behavior the most, while lack of vision of the least complex will affect behavior the least. Complexity is somewhat loosely defined as one extreme of a general dimension underlying motion, namely the degree of invariance or intrinsic organization associated with or demanded by a given movement component. Where little or no freedom of variation is tolerated the movement component is said to be complex; where much variation in the execution of a component movement may occur it is said to be relatively simple. Difficulty of movement or degree of organization may well be just as appropriate terms. In this study the experimental task was designed to make the operational component motion the most complex, followed by instrumental and reactive components.

Method

The experimental task consisted of successively transporting plastic steel-shafted pins, using a pair of thin long nose pliers, from one set of holes to another. The pins were arranged upright in a vertical line at the left edge of the $11\frac{1}{2} \times 22$ in. black workboard, which was inclined 45° and contained three $\frac{3}{8}$- and three $\frac{5}{8}$-in. "target" holes, corresponding to the two levels of precision of overall movement. An *S* grasped the specially grooved pins with the pliers and plunged them into a $\frac{1}{2}$-in. layer of soft sponge rubber beneath the six target holes in a specified order, doing all of one size in succession. The rubber pad was changed at regular intervals to prevent deterioration.

The three main components of motion in this task, then, were an *S*'s hand and arm, the pliers, and the pins which corresponded, respectively, to reactive, instrumental, and operational effects. To provide feedback of only one or combinations of these movements, a high-resolution closed-circuit television system was used. The *S*s did not look directly at the workboard, which was to their right and separated from them by a cloth screen, but instead viewed their performance on a television monitor (Conrac Model CN A8/C). A television camera (General Electric Model 4TE15B2) with 1-in. lens was mounted in front of and perpendicular to the workboard. By turning down the brightness and beam controls of the television system and elevating contrast and target abnormally the picture on the monitor was such that only white objects could be seen by *S* while black stimuli were not differentiated from the background. Accordingly, depending upon the feedback condition, *S*s wore a black or white glove and used a black or white pliers to move a set of black or white pins into the six white target holes.

Seven feedback conditions and a control were used, as defined by what *S* saw: hand-pliers-pins or in the terminology of this paper reactive-instrumental-operational (RIO), reactive-instrumental (RI), instrumental-operational (IO), reactive-operational (RO), reactive only (R), instrumental only (I), operational only (O), and a control (C) that wore no glove and viewed a normal television picture. In all cases the target holes were white and thus always seen on the monitor, as was a $\frac{1}{8}$-in. vertical white paper strip placed adjacent to the column of pins to indicate their position. Accordingly the picture on the television screen that *S* saw was entirely black except for the white target holes and white strip, plus the additional feedback provided in a particular experimental condition. (For example, in the IO condition, in addition to the holes and indicating strip, the pliers were seen as moving toward the visible pins, although *S*'s hand was invisible.) In the conditions where the pliers and/or pins were black and consequently invisible against the black background of the work board, a small segment of their outline could be detected when passed over the white strip and target holes.

Six female *S*s, paid volunteers from a nearby college, performed four trials at each of the eight conditions and two levels of precision or target hole size daily over a period of 10 days. A trial consisted of putting three pins into three holes. Order of feedback and order of hole size were randomized between and within *S*s, although each *S* received the same order of feedback on each of the 10 days. Time to completion was the dependent variable and was recorded on two clocks, one for each size of target hole, accurate to .001

min. Thus five variables—feedback, hole size, days, trials within days, and *Ss*—were represented in the analysis of variance design, and while all 3,840 scores were analyzed the results for practice are presented in terms of the days variable only.

Results

By providing visual feedback of only certain components of motion with the remaining component(s) not fed back performance was differentially affected ($p < .001$) depending upon the movement pattern that *Ss* saw. Figure 1, where average time to completion of a trial in each feedback condition is plotted against days of practice, shows that while some learning occurred at all conditions there was a general trend toward more learning at the more difficult feedback conditions (Days × Feedback, $p<.01$). The results of RIO, IO, and Control feedback had very similar effects upon performance and are thus shown as a single curve in Fig. 1. A summary of the analysis of variance of the time scores on which the above two statistical conclusions as well as subsequent statistical results are based is shown in Table 1.

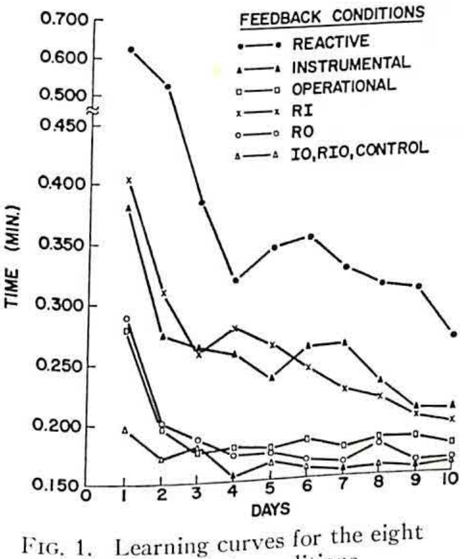


Fig. 1. Learning curves for the eight visual feedback conditions.

TABLE 1

Analysis of Variance of Time Scores (Thousands of a Minute)

Source	*df*	*MS*	*F*
Feedback (F)	7	2.4807	49.71***
Hole size (H)	1	2.8534	37.94***
Days (D)	9	.6261	6.56***
Ss	5	.4258	
F × H	7	.2100	10.19***
F × D	63	.0579	1.88**
F × *Ss*	35	.0499	
H × D	9	.0461	3.14**
H × *Ss*	5	.0752	
D × *Ss*	45	.0955	
F × H × D	63	.0151	1.22
F × H × *Ss*	35	.0206	
F × D × *Ss*	315	.0308	
H × D × *Ss*	45	.0147	
F × H × D × *Ss*	315	.0124	
Total	959		

** $p < .01$.
*** $p < .001$.

By analyzing the differences among the overall means of the eight feedback conditions it was found that those conditions where no operational feedback was provided were significantly more difficult than those in which *Ss* saw the pins. This was true both on the first day of practice and on the last day also. In addition, instrumental feedback was more important in this task than reactive feedback. Specifically, using range tests (Duncan, 1955) there were three groups of means which differed significantly from each other at the .01 level of confidence: (*a*) Control (.165 min.), RIO (.162), IO (.169), RO (.183), and O (.188); (*b*) I (.255) and RI (.257); and (*c*) R (.369). Means within any one group were not significantly different. Thus the effects of feedback fall into three classes: (*a*) those four experimental conditions providing operational feedback, regardless of additional feedback (O+); (*b*) those two conditions providing instrumental feedback but no operational feedback ($\bar{\text{O}}$I+); and (*c*) the

single condition providing neither instrumental nor operational feedback ($\overline{OI}$).

It should be noted that when all the raw-time scores were converted to log values and then analyzed the resulting curves and statistical conclusions were essentially the same as presented here.

As shown in Table 1, the overall data were also analyzed to determine the interaction between feedback and the overall precision of movement, as defined by the two sizes of target holes. Performance was naturally better ($p < .001$) on the large holes than the small holes. In addition, with practice performance became increasingly better on the large holes relative to the small holes (Size × Days, $p < .01$). Figure 2, where average time to completion on each hole size is plotted against the eight feedback conditions and summed over all scores, shows that the differences between placement of the pins in the large and small holes became increasingly greater as feedback became less helpful. That is, a significant interaction between feedback and size of hole ($p < .001$) occurred, where differences in performance between hole size became greater when no operational feedback was provided. The triple interaction of Feedback × Size × Days was not significant.

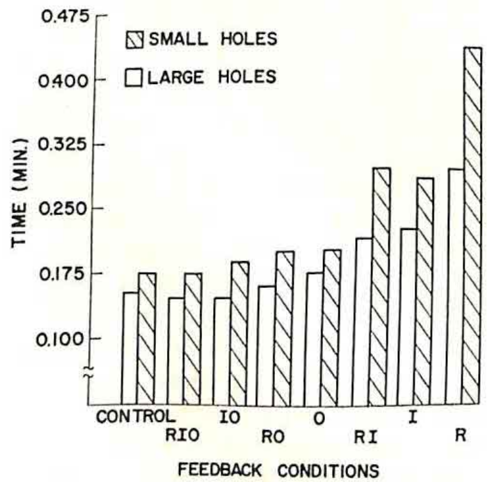


FIG. 2. Differences between overall performance on the two hole sizes as a function of feedback.

Duncan range tests were run on the 16 means of Fig. 2, as well as the corresponding means of Day 10 alone. On the overall data the 16 means fell into four groups, each group differing from each other at the .01 level of confidence. Letting lower-case characters represent performance on the small holes and upper-case represent performance on the large holes, the groups of means from worst to best were: {r}; {R, ri, i, I, RI}; {i, I, RI, ro, o}; {I, RI, ro, o, io, O, c, rio, RO, C, IO, RIO}. Means within a particular group were not significantly different. The range analysis of Day 10 showed two groups of means differing at the .01 level: {r, R, ri, i} and {R, i, ri, io, o, RI, I, c, ro, rio, O, C, RO, IO, RIO}. Overall, just the feedback condition where *S* saw only his hand (R) differed significantly as a function of hole size, while Cond. RI vs. ri and I vs. i just missed the 5% confidence level. On Day 10 no significant differences occurred within any feedback condition as a function of hole size. Considering performance on the two hole sizes as a function of feedback, the main finding was that overall time scores on the large holes did not differ for the seven easiest feedback conditions while performance on the small target holes did not differ significantly for the easiest five feedback conditions (Fig. 2). On Day 10 performance under R feedback on the small holes differed significantly from the other 15 conditions. No significant differences were observed within this latter group. By pooling the overall data to form six groups corresponding to performance on the two hole sizes for the three signifi-

cantly different feedback groups determined above ($O+$, $\bar{O}I+$, $\overline{OI}$) only the $\overline{OI}$ feedback condition differed significantly as a function of hole size.

At the end of the 10-day experiment *S*s attempted to fill the holes under two additional conditions: when they could not see their hand, pliers, or pins; and when, in addition, they could not see the holes or the thin white strip indicating the location of the pins also. These conditions were not included in the original eight because of their difficulty and they were run at the end of the experiment to determine the degree to which *S*s were still relying upon vision in this well-practiced task. In both cases performance was significantly poorer than in the main study, the completely blind condition being nearly impossible.

Discussion

The main results of this study, using a television system to control visual feedback of component movements, indicated performance was dependent both upon the components that were fed back and also the nature of the task itself. That is, the more precision demanded in the overall motion pattern, the more important was visual feedback, a fact confirming one of the predicted hypotheses. Within each level of precision the same relative order of importance of visual feedback was observed, with operational feedback affecting performance the most and reactive feedback being of least significance. In addition, an interaction occurred between feedback and level of precision, where operational feedback became even more important as the required overall movement became more precise. The rate of learning was shown to be a general derivative function of visual feedback.

The findings of this study provide evidence of the interdependency of visual perception and motion, and indicate that the effects of visual feedback of component movements are not linear, independent, or additive. Rather, for the task studied here, performance was determined specifically by the component motion fed back in the order operational-instrumental-reactive, and within this order the addition of lower-order feedback did not change performance appreciably. That is, operational feedback affected behavior the most and in the four conditions where operational feedback was provided any additional visual feedback caused no significant improvement in performance. The same conclusion is true for instrumental feedback, since no differences were found in the two conditions providing instrumental feedback but no operational feedback. Finally, the single condition providing only reactive feedback was most deleterious to behavior.

The results of this study are interpreted to support the hypothesis that the effects of visual feedback are dependent upon the complexity of the motion component involved, since the effects of feedback followed the same order as did the a priori designed complexity of component movements. Naturally a criticism of this interpretation may be raised. Although the experimental task was designed to make the operational component movement the most complex, it could be argued that this hierarchy is always the case. That this is not an entirely valid criticism, however, was shown in a study reported in Smith and Smith (1962) on simple tracing of an outline. Using three feedback conditions corresponding to RIO, RI, and O in the terminology of this paper no difference in overall performance was found.

A second criticism, related to the first, is that what has been called "complexity" is not necessarily either an apt word or descriptive of the main underlying dimension of motion. One alternative would be the degree to which each component motion is related to the completion of the task. That is, operational movement is the final step in the task and therefore the most critical. But this

is just another way of stating a fixed hierarchy in the order R-I-O, and this hierarchy has been shown to be invalid. On the other hand, the interpretation given in Smith and Smith (1962) to their many studies indicates that complexity, precision, or intrinsic organization is indeed the underlying dimension of motion.

A third criticism of this study is that hole size may be directly related only to the operational component motion and thus this independent variable did not vary the overall nature of the task as it was designed to do. If this were the case then it would be expected that the feedback of operational movement would be mainly affected by hole size. To a minor extent this is true. Based upon overall data, while hole size affected all feedback conditions the differences were larger for the three conditions where operational effects were not fed back yet this difference was significant in one condition (R) only. To answer fully these questions more research is needed wherein the complexity of each of the component motions is varied independently within the same task for several different tasks of unequal difficulty.

REFERENCES

BRIGGS, G. E., FITTS, P. M., & BAHRICK, H. P. Effects of force and amplitude cues on learning and performance in a complex tracking task. *J. exp. Psychol.*, 1957, **54**, 262–268.

CHERNIKOFF, R., BIRMINGHAM, H. P., & TAYLOR, F. V. A comparison of pursuit and compensatory tracking under conditions of aiding and no aiding. *J. exp. Psychol.*, 1955, **49**, 55–59.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

GOULD, J. D. Stereoscopic television pursuit tracking. *J. appl. Psychol.*, 1964, **48**, 369–377.

GRETHER, W. F. Efficiency of several types of control movements in performance of a single compensatory pursuit task. In P. M. Fitts (Ed.), *Psychological research on equipment design.* Washington, D. C.: United States Government Printing Office, 1947.

HOWLAND, D., & NOBLE, M. E. The effect of physical constants of a control on tracking performance. *J. exp. Psychol.*, 1953, **46**, 353–360.

SMITH, K. U. *Behavior organization and work.* Madison, Wisconsin: Univer. Wisconsin Press, 1962.

SMITH, K. U., & SMITH, W. M. *Perception and motion.* Philadelphia: Saunders, 1962.

(Received January 18, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 269–275

TRAINING AND TIMING IN THE GENERALIZATION OF A VOLUNTARY RESPONSE[1]

SHELDON H. WHITE

University of Chicago

Generalization of a voluntary response shows no increase with amounts of training between 0 and 30 trials. It is associated with short-latency responding. In view of these findings, the process of response selection in such generalization cannot easily be explained solely in terms of a habit mechanism. It is suggested that factors speeding or slowing response delivery have the effect of altering S's selection among temporally stacked response tendencies.

Under some circumstances, a human *S* instructed to give a response to a specified stimulus may show a graded tendency to deliver that response to other, similar stimuli in proportion to their similarity. Such generalization of a voluntary response was first described in a paper by Gibson (1939). It has been actively investigated during the last dozen years, after a paper by Brown, Bilodeau, and Baron (1951) introduced a widely used reaction-time technique for the study of such generalization.

The generalization of a voluntary response differs in several respects from the generalization of a classically or instrumentally conditioned response. While these differences have been commented upon by the above-mentioned authors, and by Mednick and Freedman (1960), the interrelations among the kinds of generalization found in the various paradigms are not completely clear.

Generalization of a classically or instrumentally conditioned response can only occur after the conditioning itself, after a series of training trials which produce and strengthen an association between training stimulus and response. One could, considering these paradigms, attempt to explain generalization by saying that physically similar stimuli partake of the habit of responding attached to the training stimulus, as Hull (1943) did in equations establishing generalization as a manifestation of generalized habit strength. Generalization would, in such a scheme, be expected to increase with training. In the case of instrumental conditioning, this appears to be true (Margolius, 1955). In the case of classical conditioning, it is at least partially true. Hovland (1937) concluded that the generalized GSR response increases with moderate amounts of training, decreasing somewhat with still further training.

Does the generalization of a voluntary response similarly increase with training? It is true that *E*s usually administer practice trials before testing for such generalization, but in those practice trials *S*s simply rehearse what they were told to do and what they did quite quickly and efficiently on their very first trial. It seemed worthwhile to ask whether training augments the generalization of a voluntary response, and accordingly the present studies examined the effects of varying amounts of practice upon such generalization.

[1] Study supported by a grant from the Ford Foundation in Behavioral Sciences and by United States Public Health Service Grant No. N-3639. The writer wishes to thank Angela Vellou and Paul Grim who collected data for these studies.

Another salient characteristic of the reaction-time method of studying voluntary-response generalization is an apparent dependence upon instructions emphasizing speed. Accordingly, certain steps were taken in the present studies to provide for careful examination of response latencies.

METHOD

The procedure used in these studies was a version of the reaction-time technique, with modifications made with an eye to the taking of latency measures. The *S* was asked to make one response to the training stimulus and another to any other stimulus, so that a latency measure might be obtained when an *S* made either a generalized or a nongeneralized response to a test stimulus. The usual reaction-time procedure offers *S* a respond-do not respond choice, and the latency of a response to a test stimulus can only be taken when it is a generalized response. Brown et al. (1951) reported that they had tried a two-response procedure like that used here, obtaining amounts of generalization comparable to those obtained in their more usually adopted one-response procedure.

The stimuli were a series of colored lights presented through one window, rather than the usual spatial array of lights. Reaction time lengthens to visual stimulation as it becomes peripheral, and this fact was felt to introduce an unnecessary complication into interpretation of spatial generalization data.

White noise was continuously piped to *S* after instructions were given, to partially mask out stray environmental noises. The family of reaction-time procedures is quite sensitive to auditory distractions, and it was hoped that the partial stabilization of auditory stimulation induced by the noise would improve the reliability of the reaction-time measures. (Comparisons of noise, no noise with some trial *S*s suggested that the noise level used neither speeded nor slowed simple reaction time.)

Two studies are to be reported. Experiment I compared two groups of 20 *S*s each, one group given 5 and the other group 20 trials of training before generalization tests. Experiment II compared three groups of 20 *S*s each, one group given 0 training, another given 30 trials of training, and the third given 30 trials of "control training" with a light quite dissimilar to the test colors.

Subjects

The *S*s of Exp. I were sophomore to senior undergraduates participating in the study as part of a first course in psychology. The *S*s of the second study were paid freshmen and sophomores recruited from biology classes; these *S*s had had no psychology courses.

Apparatus

The apparatus was housed in two stacked $21 \times 15 \times 11$ in. cabinets. The front face which was presented to *S* was 22 in. tall and 21 in. wide. Centered $3\frac{1}{2}$ in. from the top was a 1-in. white ready light. Centered 4 in. below the ready light was a $3\frac{1}{2}$-in. diameter circular window of flashed opal glass, through which the stimuli were presented. At bottom, centered, was a metal response panel $2\frac{1}{2}$ in. wide and 4 in. long. The *S* could rest the heel of his hand upon the table and manipulate this lever with his fingers.

In the upper cabinet, behind the stimulus window, was a semicircular array of six 75-w. projector bulbs, each connected to its own autotransformer to allow adjustments in light intensity. A hue series was arranged by placing Corning molded glass filters in front of five of the bulbs. The hues, and filters, were: (*a*) red, No. 2408; (*b*) reddish-orange, No. 2424 and 2434; (*c*) orange, No. 3307 and 3480; (*d*) yellowish-orange, No. 3486 and 3482; and (*e*) an orange-yellow, No. 3484. In these studies, Hue *a* was always the training hue, and the others were the test hues. The series was comparable to one used in a different generalization procedure (White & Spiker, 1960), except that here subjective brightnesses of the stimuli were roughly equalized.

Five preliminary *S*s were used as *O*s in method of limits comparisons where the brightness of S_1, set constant, was compared with varying brightness of each of the other hues. Ten estimates of a point of subjective equality were obtained from each *O* for each hue, and all estimates were averaged to determine a final autotransformer setting for each bulb. Reduction of current to a bulb can cause it to reach full brightness perceptibly more slowly when the bulb is turned on. Examination of the series of hues after the psychophysical adjustments indicated that the colored lights were not discriminable by differing "on-times." Both preliminary and experimental work were done in a darkened room.

The lower cabinet contained the white noise generator, and circuitry for a standard trial sequence: (*a*) with the pressing of a start button, a preset ready interval of from

½ to 4 sec. was initiated and the ready light came on; (*b*) at the end of the ready interval, the stimulus light came on, and a Standard Electric clock started; (*c*) *S*'s first response, up or down, stopped the clock, turned off the stimulus and ready lights, and registered on indicator lights in back. The trial would not begin unless *S* was holding his lever in the middle position. A few *S*s found it difficult to consistently feel this middle position. To help in such cases, a tiny indicator light was mounted at the top of the display, which lit whenever the response lever was in the middle position. After each trial, *E* recorded the time of response and whether the response had been an up or down hand movement.

Procedure

In Exp. I, all *S*s were initially instructed to lift their hand when they saw S_1, practiced this three times, and were then further instructed to press down on the lever whenever a different color appeared. Speed was repeatedly urged. There followed 5 trials of practice on S_1 for Group 5, or 20 trials for Group 20, followed without pause by a test sequence of 12 test trials interspersed among 16 presentations of S_1. Within the test trials, the four test stimuli were each given three times in a standard counterbalanced, three-cycle arrangement.

In Exp. II, *S*s in Group 30 were first instructed to respond to S_1 by lifting their hand as rapidly as possible, and did this for 30 trials. The *S*s in Group C-30 had equivalent practice with an unfiltered light. Following this practice period, the two groups received instructions identical to those of Group 0. That is, they were shown S_1 and told to lift their hand whenever seeing this color, but to press down whenever any other color appeared. Again, speed was urged. A test sequence like that of Exp. I was then completed.

RESULTS

Generalized Responses

During the test sequence, lifting the hand to S_1 was scored as correct, lifting it to a test stimulus was scored as a generalization error. Table 1 shows the percentages of such responses given by all groups. There was little difference in total generalization among groups of either study; if anything, increase of practice trials slightly decreased the incidence of generalized responses. The *S*s of Exp. I generalized somewhat more than those of Exp. II, a fact which will be discussed below.

TABLE 1

PERCENT VOLUNTARY GENERALIZATION ERRORS AS A FUNCTION OF TRIALS OF TRAINING

Group	(Correct to S_1)	Errors to Test Stimuli				
		S_2	S_3	S_4	S_5	S_2–S_5
5	(96)	83	52	42	39	49
20	(98)	70	56	41	41	47
0	(92)	72	50	28	27	44
30	(93)	52	28	30	25	34
C-30	(96)	68	33	15	18	34

Note.—$N = 20$ in each group.

Analyses of variance testing the effects of Groups and Stimuli on number of generalized responses during the test period were done for Exp. I and II separately, and for the two experiments combined. The three analyses were alike in showing highly significant effects for Stimuli, with effects of Groups and Groups × Stimuli nowhere significant.

Latency of Response

Table 2 gives the geometric mean latencies of responses during the test sequence for each of the five groups. For each *S*, every response latency was converted to log form, and four means were obtained—means for correct responses, and for errors, to the test stimuli, and means for correct responses, and errors, to the 16 presentations of S_1 interspersed among the test trials. Table 2 shows the mean of those means for each group, reconverted into seconds because it was felt the data would be more comprehensible in the common metric. The indicated statistical analyses were, of course, done with the data in log form.

It is evident from the left half of Table 2 that an error response on a

TABLE 2

GEOMETRIC MEAN LATENCIES OF CORRECT RESPONSES AND ERRORS DURING THE TEST SEQUENCE

Group	Test Stimuli		Training Stimulus		p^a	
	Correct	Error	Correct	Error	Test Stimuli Correct vs. Error	Training Stimulus Correct vs. Error
5	0.63	0.51	0.52	0.53	.01	*ns*
20	0.63	0.49	0.53	0.46	.001	*ns*
0	0.73	0.66	0.61	0.55	.02	.01
30	0.74	0.67	0.63	0.59	.05	*ns*
C-30	0.74	0.66	0.61	0.55	.01	*ns*
All *S*s	0.69	0.59	0.58	0.52	.001	.02

[a] *t* test.

test trial tended to be of significantly shorter latency than a correct response. The absolute latency difference was not large. It tended to be of about the same magnitude from the beginning to the end of the test sequence.

The fact that errors to the test stimuli were of shorter latency than correct responses might argue that impulsiveness of response helped cause such errors. Another possibility which must be dealt with, however, is that the error response—uniformly "up" for all *S*s—was physically easier, and thus quicker, than the correct response —uniformly "down" for all *S*s. Several findings argue against such an interpretation of the latency differences.

First, in these studies a significant tendency to short-latency errors was found during interspersed trials with S_1, even though opposite hand motions defined correct responses and errors on such trials. The right half of Table 2 shows this. The effect is not so regular here as on the other side of the table, but error incidence was very low on this type of trial, averaging 0.8 errors per *S* over 16 trials and estimates of error latency were accordingly much less reliable.

Second, in other research (White & Grim, 1962) a group of 20 *S*s was included in a procedure comparable to that of the present studies, except that concurrent GSR recordings were taken, and that the instructions called for reverse hand motions to those used here. For this group, error latencies to the test stimuli (0.63 sec.) were again shorter than correct response latencies (0.67). This latency difference was consistent with those found in these studies, though not statistically significant.

Third, the analysis to be discussed in the following section showed a further relationship between quick responding and errors which appeared difficult to construe as the product of a movement artifact.

Subject Speed and Errors

As noted above, *S*s of the second study generalized less than those of the first study. Since the two studies were quite comparable in most respects, about all that could reasonably be seen as a probable cause of the difference was an atmosphere effect. The *S*s of the first study, recruited in a psychology class with a lab, seemed reasonably relaxed when participating in the study. The *S*s of the second study had almost no acquaintance with psychology and were visibly more diffident.

Each S's mean log latency on his 16 S_1 trials during the testing sequence was plotted against the number of errors he had made during the test trials. Figure 1 shows this comparison; in that figure, six classes of Ss with comparable mean latencies are grouped. The decline in errors on the test trials with increasing latency on the S_1 trials was highly significant, with $F(5, 94) = 6.47$, $p < .001$. The square dots in the figure indicate mean values for all Ss in Exp. I and II; the differences in generalization in the two studies could reasonably be linked to the relationship pictured in the figure. Apparently, the atmosphere effect influenced generalization by affecting the speed at which Ss were led to pace themselves during the task.

Latency Gradients

Latency gradients of voluntary generalization have been a subject of some interest and, accordingly, those obtained in these studies are pictured in Fig. 2. Analyses of the data represented in the figure did not show significance for the effects of Stimuli or Stimuli × Correctness; however, such statistical analysis was compromised by the fact that few Ss gave both a

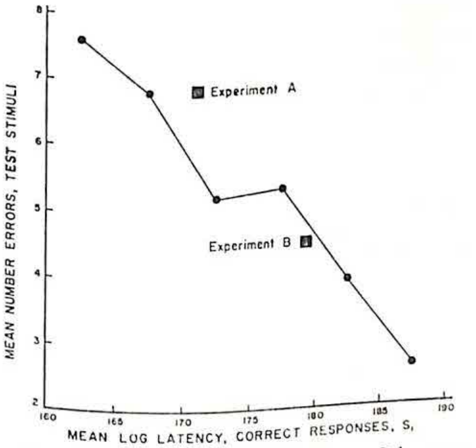


FIG. 1. Relationship between Ss' mean log latencies on interspersed S_1 trials and mean errors on test trials. (Based on 100 Ss grouped into six classes on basis of latencies.)

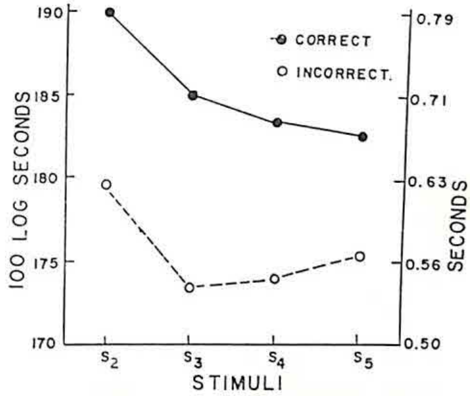


FIG. 2. Mean latencies of correct responses and errors to each test stimulus. (Based on 100 Ss.)

correct response and an error to all four test stimuli. Corresponding curves for each of the five groups were grossly comparable to their average, pictured here.

Discussion

If the generalization of a voluntary response does not increase with training, and if it is associated with short-latency responding, then an explanation of the phenomenon would apparently require some mechanism in addition to *habit* to fully describe the process of elicitation of generalized or nongeneralized responses by test stimuli. Of course, the fact that verbal instructions can call forth intended behavior must rest on habit structures, on S's previous associations between words and cues, and words and acts. Nevertheless, as the term *habit* is ordinarily understood, a habitual association between a cue and an act is not instantly installed by instructions; it strengthens with practice; it does not depend upon an appropriate response latency to take effect.

Practice has been shown to increase the generalization of one sort of voluntary response (Spiker, 1956a, 1956b). In Spiker's procedure, the response measure was a count of the number of handle pulls within a 3- or 4-sec. interval. The speed and rhythm of such bursts of responding may be increased by practice and thus, secondarily, practice could augment recorded responsiveness during general-

ization tests. However, in a reaction-time procedure the learning or smoothing of response-response sequences is probably of minimal importance, and thus training has no apparent effect.

The influence of short- or long-latency responding upon the generalization of a voluntary response suggests that the phenomenon is associated with temporal response priorities. In fact, Fig. 3 exactly illustrates a temporal stacking of the two competing responses in the generalization test trials of the present studies. Perhaps something like the following is involved in the process of response selection in the present procedure:

With the first presentation of S_1 to him, *S* stores an approximate model of the stimulus—this is, of course, self-evident in that *S*s can recognize something they have only seen once. Similarly, *S* stores two response sets as a result of instructions—R_1 if S_1; R_2 if anything else. It seems plausible that the instructions establish the responses not as simultaneously competing, but in a priority order—"R_1—and if not, R_2."

Two delays of response are apparently involved in the subsequent use of these stored specifications, what might be called a general and a contingent delay. The majority of responses during testing

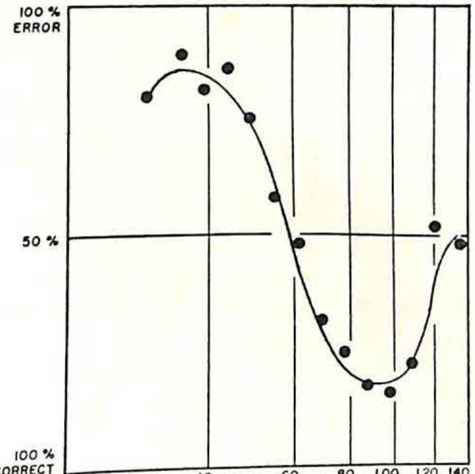


FIG. 3. Proportions of correct or incorrect responses delivered at successive (logarithmically equal) intervals after stimulus onset. (Pictured time-space includes almost all of 1,200 test-trial responses.)

—correct responses to S_1, and errors to the test stimuli—were delivered at a mean time of .58 sec. In the study, these were all "up"—or, in the terminology adopted for discussion, R_1—responses. This mean time for the first-available response is considerably longer than the usual simple RT, and consequently it seems reasonable to suppose that *S* is operating under a general delay. Such a delay, common to all disjunctive RT procedures, presumably gives time for matching of the perceived stimulus against the stored model.

If a mismatch is detected, then a second, contingent delay appears to be necessary for delivery of R_2. Perhaps it is associated with the sort of novelty reaction shown in Grings' (1960) perceptual-disparity response, or the orienting reflex (Sokolov, 1960). The contingent delay is quite brief, and may simply be a consequence of the fact that response must be inhibited until R_2 is available.

One must assume that a response can be launched before these delays complete themselves, and a response so launched will reflect whatever response is available at that instant. Any long-term condition (such as a speed set) or short-term condition (such as startle) which forces out a response before a contingent delay will tend to induce generalization on a test trial.

Occasionally, a response may be forced out before even general delay is completed—in which case, response selection would tend to be random. This, it is assumed, is the genesis of the few short-latency errors to S_1 which occurred; such responses were the very fastest responses given in the testing sequence.

The above explanation of the response-selection process in the generalization of a voluntary response is necessarily sketchy. It deals with choice of response to the generic test stimulus, leaving aside important unanswered questions as to how similarity between training and test stimuli influences the incidence of errors. It does, however, offer some grounds for understanding of *S*'s choice between speed or accuracy in rational terms and, in so doing, it suggests tie-ins among a number of significant aspects of the generalization literature.

It accords with the fact that instructions emphasizing speed are so important to voluntary generalization (Brown et al., 1951), and that impulsive *S*s tend to be high generalizers (Lacey & Lacey, 1958).

It is in line with the fact that associative interference errors, logically much like generalization errors, are associated with short-latency responding (White, Spiker, & Holton, 1960).

It allows a reasonable explanation of Mednick's (1955) finding that organic patients generalize less than normals. While it seems hard to believe that heterogeneous brain damage leads to more accurate stimulus differentiation, it would be quite in line with what is known of organic patients (Goldstein, 1947) to suggest that they adopt a more hesitant, slower speed set on a generalization task.

The fact that schizophrenics generalize more than normals (Knopf & Fager, 1959; Mednick, 1955) could under this view be linked with "microgenetic" processes. Recently, theory (Flavell & Draguns, 1957) and rather impressive evidence (Werner, 1957) have been offered to suggest that schizophrenic reactions resemble impulsive reactions usually inhibited by normals.

Finally, the findings of Thompson (1962) appear to be in line with the present sort of reasoning. He has reported that removal of auditory cortex in cats led to high levels of generalization, apparently by creating a loss of ability to form response inhibition.

REFERENCES

Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.

Flavell, J. H., & Draguns, J. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54**, 197–217.

Gibson, E. J. Sensory generalization with voluntary responses. *J. exp. Psychol.*, 1939, **24**, 237–253.

Grings, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **27**, 243–252.

Goldstein, K. *Human nature in the light of psychopathology.* Cambridge, Mass.: Harvard Univer. Press, 1947.

Hovland, C. I. The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, **21**, 47–62.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Knopf, I. J., & Fager, R. E. Differences in gradients of stimulus generalization as a function of psychiatric disorder. *J. abnorm. soc. Psychol.*, 1959, **59**, 73–76.

Lacey, J. I., & Lacey, B. C. The relationship of resting autonomic activity to motor impulsivity. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1958, **36**, 144–209.

Margolius, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.

Mednick, S. A. Distortions in gradients of stimulus generalization related to cortical brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **51**, 536–542.

Mednick, S. A., & Freedman, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

Sokolov, E. N. Neuronal models and the orienting reflex. In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* Madison, N. J.: Macy Foundation, 1960.

Spiker, C. C. The effects of number of reinforcements on the strength of a generalized instrumental response. *Child Develpm.*, 1956, **27**, 37–44. (a)

Spiker, C. C. The stimulus generalization gradient as a function of the intensity of stimulus lights. *Child Develpm.*, 1956, **27**, 85–98. (b)

Thompson, R. F. Role of the cerebral cortex in stimulus generalization. *J. comp. physiol. Psychol.*, 1962, **55**, 279–287.

Werner, H. The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development.* Minneapolis: Univer. Minnesota Press, 1957.

White, S. H., & Grim, P. F. Investigations of a voluntary generalization paradigm. *Amer. Psychologist*, 1962, **17**, 378.

White, S. H., & Spiker, C. C. The effect of a variable conditioned stimulus upon the generalization of an instrumental response. *Child Develpm.*, 1960, **31**, 313–320.

White, S. H., Spiker, C. C., & Holton, R. B. Associative transfer as shown by response speeds in motor paired associate learning. *Child Develpm.*, 1960, **31**, 609-616.

(Received January 20, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 276–281

EFFECTS OF STIMULUS CHANGE UPON THE GSR AND REACTION TIME[1]

PAUL F. GRIM AND SHELDON H. WHITE

University of Chicago

60 Ss were given 16 RT trials using a colored light as stimulus and then, without warning, lights differing in color but not in intensity were presented. Augmented GSR reactions occurred as a monotonic function of amount of change. Lengthened RTs were also observed; the amount of such lengthening did not appear to be simply related either to evoked GSR or to amount of stimulus change.

When *S*s are presented with a series of conditioning trials, and neither the stimulus nor the reinforcement undergoes a discernible change, *S* appears to become habituated to the constant characteristics of his repeated experiences. If a change then occurs in either the stimulus or the reinforcement, a drop in skin resistance occurs; this GSR reaction has been referred to as a *perceptual disparity response.* The magnitude of the perceptual disparity response is a function of the number and proportion of previous experiences which have been consistent with one another, and of the amount and direction of the stimulus change involved (Grings, 1960).

Kimmel (1960), using auditory stimuli, studied the effect of changes in stimulus intensity upon the perceptual disparity response, and found that amount of GSR reaction related monotonically to the extent of stimulus change. The present study sought to determine the extent of perceptual disparity responses to a series of colors for which generalization data had previously been obtained. The intent of the study was to look for possible points of connection between the perceptual disparity response and generalization. Recordings of reaction time were made, to see if the perceptual disparity response is accompanied by any change in timing of response to a stimulus, since response timing appears to be a significant influence on the generalization of a voluntary response (White, 1965).

[1] Study supported by a grant from the Ford Foundation in Behavioral Sciences and by United States Public Health Service Grant No. M-3639. The principal responsibility for the interpretations made of these data rest with the second author.

METHOD

Apparatus.—The basic apparatus has been described in some detail in White (1965). Essentially, it consisted of two stacked cabinets, the top one containing lights and filters for the presentation of a series of illuminated colors, the bottom cabinet presenting to *S* a response lever, and containing switching and timing circuitry for a simple RT procedure.

Continuous GSR recordings were obtained using a Fels dermohmmeter, Model 22A, connected to a 1-ma. Esterline-Angus pen recorder. Subject current was 70 μa. Electrodes were circular $\frac{13}{16}$-in. zinc disks, set in a $\frac{1}{8}$-in. deep Plexiglas cup filled with electrode jelly supplied by the instrument manufacturer (Yellow Springs Instrument Company, Incorporated, Yellow Springs, Ohio).

Stimuli.—Stimuli were presented through a $3\frac{1}{2}$-in. diameter circular window of flashed opal glass. Behind the window was a semicircular array of five 75-w. projector bulbs, each connected to its own autotransformer to allow adjustments in light intensity. A hue series was arranged by placing Corning molded glass filters in front of these bulbs. The hues, and filters, were: (*a*) red, No. 2408; (*b*) reddish-orange, No. 2424 and 2434; (*c*) orange, No. 3307 and 3480; (*d*) yellowish-orange, No. 3486 and 3482; and (*e*) an orange-yellow, No. 3484. This series, or a

closely similar series, had yielded monotonic generalization gradients in a number of previous studies (White, 1965; White & Grim, 1962; White & Spiker, 1960).

An attempt was made to roughly equate subjective brightnesses of the stimuli through the following procedure. Five preliminary *S*s were used as *O*s in method of limits comparisons where the brightness of S_1, set constant, was compared with varying brightnesses of each of the other hues. Ten estimates of a point of subjective equality were obtained from each *O* for each hue, and all estimates were averaged to determine a final autotransformer setting for each bulb. Both preliminary and experimental work were done in a darkened room.

Subjects.—Sixty undergraduates at the University of Chicago served as *S*s. Twenty-three were from a psychology course, and fulfilled a laboratory requirement by participating in the experiment; these *S*s were given only one "change" trial and will be termed Group A. Thirty-seven *S*s were paid volunteers recruited from an undergraduate biology course; these *S*s were given four successive change trials and will be termed Group B. Except for their differing numbers of change trials, both groups experienced identical procedures.

Procedure.—Electrodes were attached to the *S*s 20 min. prior to the experiment, in order to allow time for the electrode jelly to hydrate the skin. Both electrodes were placed on the palm of the left hand, one just under the index finger and the other a maximal distance away on the heel of the hand. Plexiglas spring clamps held the electrodes in position.

After the preliminary period, spent in a waiting room, *S* was brought into the experimental room and seated before a table holding the reaction-time apparatus. He was asked to use his free right hand to press the lever while avoiding tensing or moving the left hand, which rested on the table. A set of written instructions explained the task. The *S*s were instructed to maintain the response lever in a middle position until the window lighted, and then to press down on the lever as quickly as possible.

The *S*s were then given 16 habituation trials to S_1, red stimulus. On the seventeenth trial, without pause or warning, one of the other four stimuli was presented. Fifteen *S*s received each one of the four change stimuli on the seventeenth trial. On Trials 18–20, the 23 *S*s of Group A returned to S_1 again, while the 37 *S*s of Group B continued to receive the change stimulus assigned to them on Trial 17.

On each trial, latency of response was recorded on a Standard Electric timer to hundredths of a second and, for analytic purposes, such latencies were later converted to logs. A GSR reaction was scored if a drop in resistance occurred within 1–3 sec. after stimulus presentation; the resistance readings were converted to log conductance change measures.

Consideration of possible GSR artifacts.—The procedure and recordings offered the possibility of several artifacts in the GSR records—due to scale changes, polarization, and movement—which, though they were not felt to be critical, do require some discussion in the context of a description of method.

The Fels dermohmmeter maintains both range and sensitivity by an automatic scale change feature, which causes a full-scale return sweep of the pen whenever it approaches maximum excursion. About 6% of the time, an automatic scale change of the dermohmmeter occurred immediately after a trial, and the consequent full-scale sweep of the galvanometer pen made it impossible to score a GSR reaction for that trial. This introduction of unscorable trials was, unfortunately, not random but tended to obscure more large than small GSR reactions—the larger reactions being more likely to bring the pen to the point of scale change. The effects of such an artifact were judged to be essentially neutral for experimental purposes, since it would tend to reduce both treatment and error GSR variance.

Electrode polarization, always a factor to be considered in GSR work, probably had little influence on the results of this study. The actual recording went on for only 10 min. or so, too brief a time for much polarization to occur. (Occasionally, the two zinc discs were touched together just after removal from an *S*, and little or no voltage differences were observed on the instrument's galvanometer.) Even had more polarization occurred, it is dubious whether it could have significantly altered a study devoted, as this one, to resistance change rather than basal resistance data.

Finally, the fact that one of *S*'s hands was working while the other held the GSR electrodes opened the door to the possibility that the GSR records contained a component of movement artifacts: that is, that the GSR pen might have at times acted as a stabilimeter recorder as well as a resistance recorder. It was felt that movement artifacts did not

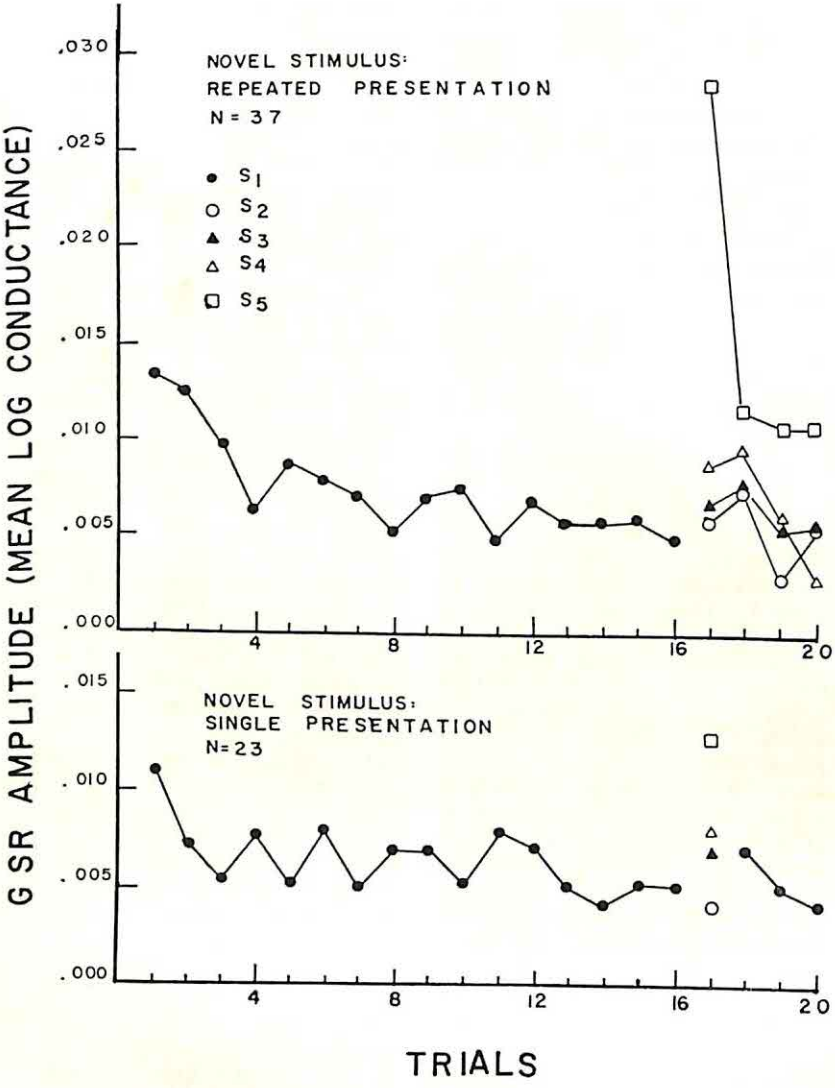


FIG. 1. Mean GSR reaction (mean log conductance change) on each trial of the study. (Group A results are pictured in the bottom part of the figure, Group B results above.)

significantly influence the obtained data because; (*a*) pretesting suggested that the electrodes on the left hand were not very sensitive to movements of the right hand; (*b*) when movement artifacts were observed, they gave a characteristic sharply swinging pen record—and such were rarely seen in the experimental records; and (*c*) the movement responses which might induce an artifact in the study came about .5 sec. after stimulus onset, while GSR reactions were scored in the interval from 1 to 3 sec. after onset.

RESULTS

Perceptual disparity responses.—Figure 1 pictures mean GSR reactions for each trial of the experiment. The data for Group A, which received only one change trial, are pictured at the bottom of the figure, the data for Group B, with four change trials, at the top.

GSR responsiveness, high during

the first few habituation trials, settled down to a low and fairly steady level within a few trials. The decline in mean GSR reaction from Trial 1 to Trial 16 was highly significant, with $t\ (48) = 4.60$, $p < .001$. (As with some other GSR comparisons to be reported, less than the total sample N were involved because of unscorable GSR trials, as noted above.)

Trial 17, the first change trial, induced a markedly elevated GSR reaction, significantly different from the reaction on Trial 16. Here $t\ (57) = 2.99$, $p < .01$. The size of the reaction was proportional to the amount of stimulus change (Fig. 2). A simple analysis of variance, comparing seventeenth-trial GSRs to S_2, S_3, S_4, and S_5 yielded an $F\ (3, 56)$ of 6.58, significant at the .005 level.

When S returned to S_1 after the change trial, his GSR reactivity immediately returned to normal range. If S remained with his change stimulus, the return to normal range appeared to be delayed by a single trial (Table 1).

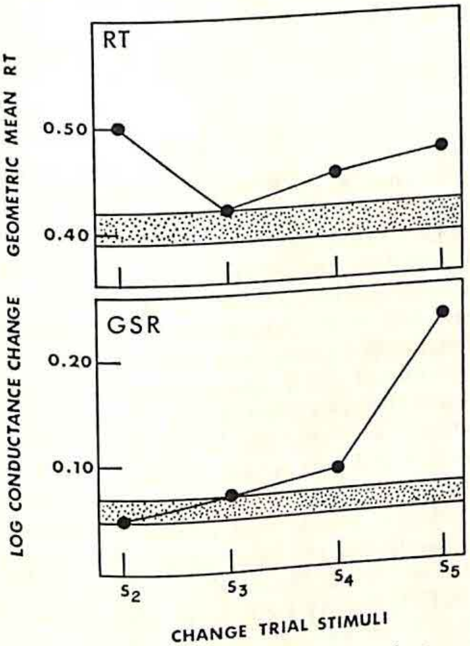


FIG. 2. Mean GSR log conductance change and geometric mean RT on Trial 17, the first change trial, for subgroups of Ss assigned to the four change stimuli. (The stippled zones indicate the range of these response measures over previous habituation trials 6–16.)

TABLE 1

LATENCY OF RESPONSE AND GSR REACTIONS DURING HABITUATION AND CHANGE TRIALS

		Habituation	Change Trials			
		Range, Trials 6–16	17	18	19	20
Mean GSR Log Conductance Change						
A	23	.005–.007	.011	.007	.006	.007
B	37			.010	.006	.007
Geometric Mean Reaction Time						
A	23	0.39–0.42	0.46	0.43	0.42	0.41
B	37			0.45	0.44	0.44

Reaction-time data.—The trends observed in the reaction-time data were grossly similar to those found in the GSR data. Reaction time quickened from Trial 1 to Trial 16, with $t\ (59) = 3.71$, $p < .001$. The extent of the change was from a geometric mean time of 0.46 sec. to 0.41 sec. On the first change trial, a lengthened RT was observed, differing significantly from the RT measure on Trial 16, with $t\ (58) = 3.59$, $p < .001$. The return to normal range was rapid for Group A, somewhat slower for Group B; on Trial 20, Group B was still significantly slower than it had been on Trial 16, with $t\ (36) = 2.08$, $p < .05$ (Table 1).

Figure 2 pictures the lengthening of RT on the first change trial. The differences in RT among the test stimuli were not supported by statistical tests. An analysis of variance comparing mean RTs to S_2, S_3, S_4, and S_5 on Trial 17 was not significant,

nor was another analysis comparing changes in RT from Trial 16 to Trial 17 for subgroups of *S*s assigned to the different change stimuli.

Correlations of RT and GSR.—The first few habituation trials were marked by long RTs and large GSR reactions. These two response characteristics appeared during the first change trial as well. Analyses were therefore undertaken to see whether RT and GSR measures tended to be quantitatively correlated.

The first such analysis looked for an association between an *S*'s average levels of RT and GSR response. It was felt necessary for this analysis to exclude *S*s who rarely gave GSR responses. Accordingly, 49 of the 60 *S*s were selected out as "active" on the GSR during habituation, on grounds that these *S*s gave a nonzero GSR on more than half of their scorable habituation trials. Each active *S*'s average RT over his habituation trials was correlated against his average GSR reaction over those trials. The correlation was .28, significant at the .025 level, suggesting that individuals who tended to be more reactive on the GSR were to some extent also slower in RT.

Would, then, moment-to-moment variations in a given individual's GSR responsiveness show a similar relationship to his variations in speed of response? For each of the 49 active *S*s, an individual correlation coefficient was computed, relating GSR and RT over his 16 habituation trials. The resulting coefficients varied widely from individual to individual (Table 2). Goodness-of-fit tests, using theoretical distributions of r calculated by Fisher's z (Walker & Lev, 1953), showed that such a distribution of r's would be unlikely to occur through sampling error from a rho of .28 ($\chi^2 = 53.56$, $p < .001$). However, the obtained distribution was not inconsistent with sampling error around a rho of 0. There was, then, no evidence for conclusion that a given *S*'s RT tended to vary from trial to trial with his GSR.

Finally, a correlation was run comparing *S*s' GSR and RT measures on Trial 17, and involving the total of 59 *S*s for whom both measures were available on this trial. The correlation was .14, not significantly different from 0.

TABLE 2

FREQUENCY DISTRIBUTION OF 49 INDIVIDUAL CORRELATION COEFFICIENTS RELATING GSR AND LOG RT DURING HABITUATION TRIALS

	Value of Correlation			
	−.55 to −.20	−.19 to .07	.08 to .34	.35 to 81
Observed	12	14	11	12
Expected, if p = .28	2	9	18	20

DISCUSSION

Kimmel's (1960) study, and the present one, have shown that the GSR is responsive in a graded fashion to changes in quality and intensity of stimulation. As Allen, Hill, and Wickens (1963) have pointed out, this fact has significant implications for the interpretation of studies of GSR generalization. Unless most response systems are as sensitive to stimulus change as is the GSR, which seems dubious, it is unlikely that GSR generalization can be well representative of the universe of generalizing behavior. This is a consideration of some importance, since Hovland's (1937a, 1937b, 1937c, 1937d) GSR generalization work has been a classic and important source of theory and data concerning stimulus generalization.

GSR data may be atypical. For example, comparison of studies of intensity generalization (White, 1962) shows that Hovland's generalization curves are un-

like all others reported in that, testing towards louder tones, he found response strength ascending away from the training stimulus. This effect, sometimes attributed to energizing properties of CS intensity, might involve also the responsiveness of the GSR to stimulus change.

Recent data have suggested, in the case of voluntary response generalization, that generalization errors and correct responses are "ready" at different intervals of time after stimulus onset. Short-latency responses are apt to be generalization errors, while some sort of delay is associated with the delivery of a correct response (White, 1965).

In the present study, stimulus change caused a lengthening of response latency by about .04 sec. Five groups of generalization *S*s (White, 1965) had delay associated with a correct response averaging .07, .07, .08, .12, and .14 sec. (comparing geometric mean latencies of correct responses and errors to generalization test stimuli). The delay induced by the perceptual disparity reaction is a little smaller, but of the same order of magnitude, as the delay involved in delivery of a correct response in a generalization test.

However, it seems unlikely that the perceptual disparity process, as it is now known, could act as the perfect complement to the generalization process. The perceptual disparity response is short-lived; it seems difficult to see how it could produce the gradual long-term decrements in generalized response which have been consistently found with repeated testing. In addition, while the GSR component of the disparity response is monotonically related to stimulus differences, the amount of RT lengthening is neither appreciably correlated with the GSR, nor does it appear to be monotonically related to extent of stimulus change.

Possibly, the perceptual disparity reaction may be a trigger process which sets into motion the antigeneralization process or processes which cause the decline of generalized responding with repeated experience with the test stimuli.

REFERENCES

Allen, C. K., Hill, F. A., & Wickens, D. D. The orienting reflex as a function of the interstimulus interval of compound stimuli. *J. exp. Psychol.*, 1963, **65**, 309–316.

Grings, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **67**, 243–252.

Hovland, C. I. The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, **21**, 261–276. (a)

Hovland, C. I. The generalization of conditioned responses: III. The extinction, spontaneous recovery, and disinhibition of conditioned and of generalized responses. *J. exp. Psychol.*, 1937, **21**, 47–62. (b)

Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148. (c)

Hovland, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291. (d)

Kimmel, H. D. The relationship between direction and amount of stimulus change and amount of perceptual disparity response. *J. exp. Psychol.*, 1960, **59**, 68–72.

Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

White, S. H. A note on intensity generalization and prothetic scaling. *Psychol. Rev.*, 1962, **69**, 149–155.

White, S. H. Training and timing in the generalization of a voluntary response. *J. exp. Psychol.*, 1965, **69**, 269–275.

White, S. H., & Grim, P. F. Investigations of a voluntary generalization paradigm. *Amer. Psychologist*, 1962, **17**, 378.

White, S. H., & Spiker, C. C. The effect of a variable conditioned stimulus upon the generalization of an instrumental response. *Child Develpm.*, 1960, **31**, 313–320.

(Received January 20, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 282–286

EFFECT OF PRIOR KNOWLEDGE OF THE STIMULUS ON WORD-RECOGNITION PROCESSES [1]

RALPH NORMAN HABER

Yale University

Giving S knowledge of the stimulus word immediately prior to its exposure increased the probability of S being able to perceive all of the letters of the word ($p < .001$). While a difference in probability of perception of the letters was found for rare as compared to frequent words ($p < .001$), this difference completely disappeared when S had prior knowledge of the word. Both of these findings seemed consistent with a response interpretation of word-recognition processes. However, giving S repeated exposures of the word increased the probability of seeing the letters, regardless of whether he had prior knowledge of the word, a result interpreted as quite inconsistent with response processes. Further, examination of Ss' reports showed that they were perceiving letters, not making guesses about the word, and that the percept of the letters gradually increased in clarity, quite independently of whether they knew the word. The similarity of these results and conclusions to Hebb's notions of the development of cell assemblies and phase sequences was pointed out.

An interpretation of word-recognition thresholds based on response processes (e.g., Eriksen, 1958; Pierce, 1963) should predict that if *S* had complete and exact information about the stimulus word, just prior to its presentation, all variation introduced by the probability of occurrence of the word would disappear. This prediction is based on the assumption that if *S* has full prior knowledge of the stimulus, there can be no further variation introduced by differential probabilities of having response items available, or of testing the correct "hypotheses" about the stimulus. The present experiment attempted to test several implications of this prediction.

The *S*s were shown both rare and frequent English words, with the number of exposures varied for each word, but not their duration or intensity. Half of the words were exposed, prior to their first flash, for 5 sec., and *S* had to spell the word to *E*. In this way, for half of the words, regardless of their frequency in print, the response probability of *S* being able to select the correct word from all of the possible words in his vocabulary was 1.00, since he knew the word exactly.

It was expected that the words for which *S*s had prior knowledge would have a higher probability of having all of their letters perceived. Further, it was also expected that there would be no difference in probability of perceiving all letters between the rare and frequent words when *S* had prior knowledge of the word. Without such prior knowledge, a difference would be found, favoring the words that had appeared more frequently in print. Both of these predictions follow from a response interpretation of recognition thresholds.

A more difficult prediction concerns

[1] This research was partially supported by a grant from the United States Public Health Service, MH-03244. Deep appreciation is acknowledged to Maurice Hershenson, whose participation with the author on earlier studies made this one possible. Special thanks are also given to Martha Breen for her assistance in the collection and analyses of the data.

the effects of repeated exposures of the word. Haber and Hershenson (1965) and Hershenson and Haber (1965) both found that the probability of perceiving all letters of a word increased markedly with repeated exposures, even though there had been no increases in either duration or intensity of the stimulus. Since their *S*s were reporting whether they saw the letters of the words, rather than whether they knew or could guess the word, they argued that the repeated exposures were affecting the development of a percept or image, rather than response probability or availability. Therefore, a "perceptual" interpretation would predict that even when *S*s have prior knowledge of the word, there will still be a gradual increase in the probability of perceiving all of the letters with repeated exposures of the word. A response interpretation would have to predict that repeated exposures should have no effect on recognition, since *S*s know what the word is.

Method

Each *S* was shown 576 seven-letter, three-syllable English words, divided into those rare and frequent. The rare words were selected from the rarest third of those appearing in the Thorndike and Lorge (1944) summary word count (G) and from words culled from Webster's Unabridged Dictionary (1939) that did not appear in the Thorndike-Lorge lists at all. The frequent words were all above the median on the summary count in the Thorndike-Lorge lists, and represent virtually the entire population of frequent English words with this structure.

These words were randomly divided into nine lists of 64 words, each list to be shown on one of the nine experimental sessions of the experiment. Each list had an equal number of randomly ordered rare and frequent words.

The apparatus and viewing arrangements were the same as in the two previous studies (Haber & Hershenson, 1965; Hershenson & Haber, 1965). The stimuli were presented in one channel of a three-channel mirror tachistoscope (Scientific Prototype Manufacturing Corporation, Model D). A second channel, serving as background, was always lighted, and contained two faint lines for fixation boundaries. The reflectance measured at the eyepiece with a Macbeth illuminometer was 10 footlamberts (ftL.) for the background and 18 ftL. for the stimulus on the background.

The *S* was always prepared for the flashes, since he initiated each trial by pressing a button to trigger the tachistoscope.

Procedure.—Each word was assigned one of two duration values (high or low), one of eight exposure trial numbers (1, 2, 3, 4, 5, 10, 15, or 25), representing the number of times it would be flashed, and one of two conditions of prior knowledge (none or complete). Duration was never changed during the presentation of a word, regardless of the number of exposures. The complete prior knowledge condition was provided by exposing the word for 5 sec. in the channel before the first trial and requiring *S* to spell out the word to *E*. The order of the frequency of the words was random, as were the number of exposure trials, and the duration of the exposures. However, every odd word was exposed first for prior knowledge. Within each session of 64 words, one word was shown for each of the 64 possible combinations of experimental variables.

Each *S* was given three practice sessions before the nine experimental sessions. These practice sessions included only frequent words, comparable to the frequent ones used in the latter session, with no prior knowledge given for any of them. The duration at which these practice words were exposed was systematically varied for each *S*, so that two durations could be determined for use in the latter sessions. The lower duration was one where *S* never identified all seven letters on the first flash, but occasionally did so when the duration was 1 or 2 msec. higher. Once this lower duration was determined, the higher duration was set as 5 msec. above it.

For each flash, in both the practice and experimental sessions, *S* reported the letters and their respective positions he was certain he perceived. The *S* was required to report the letters, even when he could identify the word. All analyses were based on the perception of letters, not of words. This maximizes reports based on what *S* saw, rather than on what he thought he saw. The *S* was scored as having perceived a word if he correctly reported seeing all of the seven letters on the last of the exposures given for that word. The *S* was given no feedback on his accuracy for the words without prior exposure. Further, *S* did not know at the time he was making his report for any trial whether there would be further trials for that same word.

This unpredictability of further exposures was stressed to *S*, since otherwise *S* could withhold his reports until he was more certain. The *S* always knew when a new word was to be presented. The *S* was given extensive instructions regarding the importance of reporting letters rather than words, and he was reminded of them when necessary.

Subjects.—Sixteen Yale undergraduates each served 12 hr. as *Ss*. They were tested individually, and had never been in a perception experiment before.

RESULTS

Figure 1 presents the results for the number of exposures (trials), duration of exposure, frequency of words, and prior knowledge. Since there appeared to be no interactions with trials, each *S*'s scores were summed across trials, and the means of those summed scores are indicated by the points on the right of the figure. For each duration, both prior knowledge means are significantly above both no-prior knowledge means. While this is more true for the low duration, $t\ (15) = 10.61$, $p < .001$, than for the high duration, $t\ (15) = 4.10$, $p < .001$,

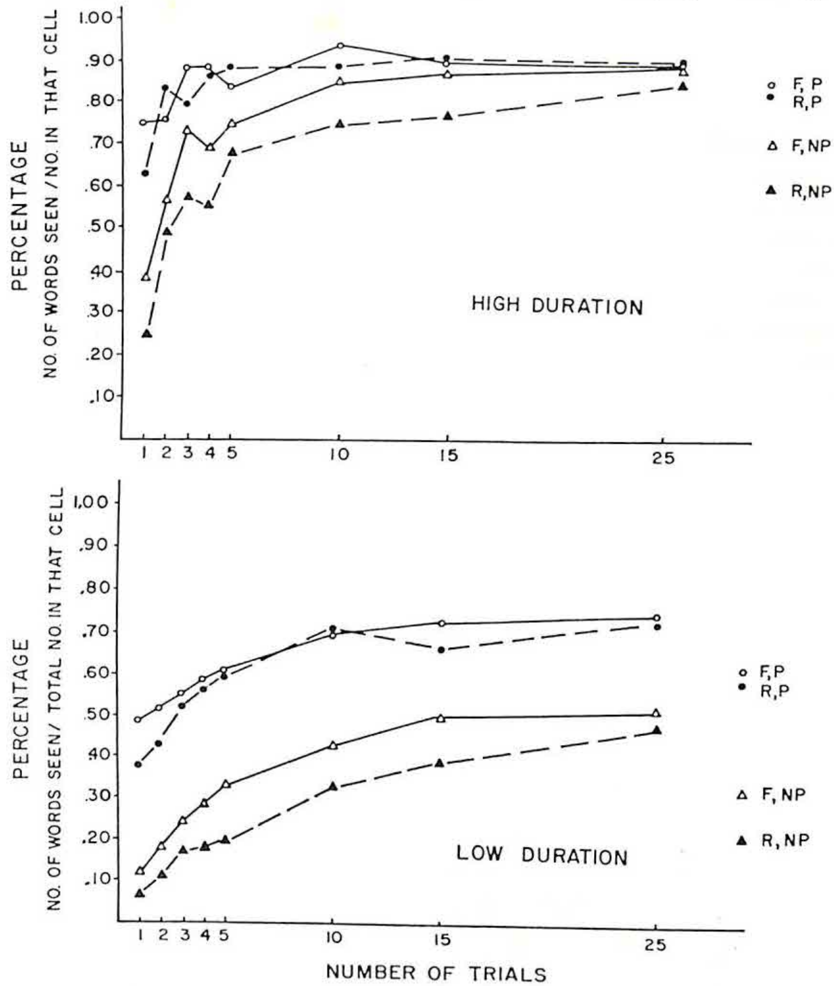


FIG. 1. Probability of perceiving all the letters of a word as a function of repeated exposure trials, frequency of the words, prior knowledge of the words, and duration of exposure. (Frequency is indicated by frequent [F] or rare [R]; prior knowledge by P and no prior knowledge by NP. The points to the right of the curves are the means summed over the total number of trials, and are the basis for all of the *t* tests.)

a difference between durations that is also significant beyond .001, the convergence of the curves for the high-duration words was more likely to be due to a ceiling, which was clearly reached much earlier for the prior knowledge words. In any event, it is clear that giving S prior knowledge of the word increases the probability of his perceiving all of its letters.

For both durations, the difference between rare and frequent words was significant only when S had no prior knowledge—Low D: $t\,(15) = 6.73$, $p < .001$; High D: $t\,(15) = 3.47$, $p < .001$. Both t tests yield values less than 1.00 between rare and frequent when prior knowledge had been given. Thus, the second prediction was supported, that prior knowledge obliterates the difference between rare and frequent words.

The findings regarding the effects of trials clearly support the "perceptual" rather than the response interpretation. This was seen most clearly for the low duration, where there was no ceiling. Here, all four curves showed nearly identical effects of trials, regardless of the frequency of the words or the prior knowledge of S. The same tendency is apparent for the high-duration curves, even though the prior knowledge words reach the ceiling first. Consequently, even when S knows exactly what the stimulus will be, his ability to see all of the letters grows gradually with repeated exposures, just as it does when he does not know the stimulus.

For purposes of comparison with the two previous studies, least-squares solutions were obtained for each of the eight curves. The identical function was obtained for each of the eight, which was also the same function found repeatedly in the previous experiments. This was

$$P_n = A - q_1/n$$

where

P_n is the probability of perceiving all seven letters by the nth trial.

A is the asymptotic probability obtained after 25 trials.

q_1 is the probability of *not* perceiving all seven letters on the first trial.

n is the number of trials.

Each least-squares fit indicated that the exponent of n did not differ from 1.00 except by sampling error, but was significantly different ($p < .001$) from zero (a formal test of the significance of the effects of trials).

Discussion

These results indicate that the effects of word frequency are probably mediated by response processes, since these effects are removed when a principal source of variance due to response processes is also removed. However, such response processes probably do not account for the more basic perceptual behavior itself. While it seems intuitively reasonable that knowing what the stimulus will be should make it more easily perceived, this makes sense only if perceiving implies the ability to make some kind of probabilistic response of tentative recognition or identification. However, the rather slow growth in the actual ability of S to see the letters of the word, even when he knows exactly what the word is, does not seem to be a change in probabilistic responses.

This is most clearly illustrated by examining the nature of Ss' reports about their percepts. On the first flash of a word, S generally reported seeing nothing at all—no letters or pieces of letters. After several flashes, first pieces and then whole letters would be visible, and after a few more presentations, all of the letters would be perfectly plain and clear. The Ss, of course, could correctly guess the word well before all of its letters were perceived, but even after they had guessed it, there was a continued gradual development of the percept of the rest of the letters. These percepts became increasingly clear and distinct, so that reports after a number of presentations

were not guesses or "hypotheses," but were based on quite an explicit and unambiguous image of the word. Thus, while it is possible that response processes might be controlling the guessing of the word, especially from images without all of the letters present, it seems quite unlikely that response processes are also controlling the gradual development of the percept of the word.

Another line of evidence supports this conclusion. The effect of trials, that is the slow growth in the probability of being able to perceive all letters as a function of repeated presentations, has now been found in three different experiments, under a variety of conditions. For example, this effect holds for frequent English words (this study, Haber & Hershenson, 1965, Hershenson & Haber, 1965), for rare English words (this study), for totally unfamiliar nonsense words (Hershenson & Haber, 1965), and words for which *S* knew their exact content immediately prior to exposure (this study). Further, the mathematical equation specifying the function of exposure trials to perceiving the words has been the same for each experiment and for each condition within each experiment. Thus, it would seem that *S*s' knowledge or past experience with the words (the stuff of which response processes would be made) is fairly irrelevant to the development of a percept.

These three experiments, even taken together, do not suggest what mechanism would account for the growth of a percept. However, such a growth is remarkably similar to Hebb's (1949) discussion of the growth of a cell assembly, and the organization of cell assemblies into phase sequences. If this analogy is correct, then the development of such assemblies could be considered independently of previously established assemblies. That is, the *speed* of associating the components of an assembly would be independent of whether such an assembly had previously been established, even though the number of components with which one begins might be determined by the existence of previous assemblies, as well as other variables. In this sense, the intensity or duration of the stimulus, and the amount of *S*'s prior knowledge or experience with that stimulus, would determine how many components are available at the beginning of the development of an assembly, under some kind of assumption that the more elements available, the faster or more completely the assembly develops. These variables might also determine the maximum degree of development, such that if the duration or intensity were too low, or the previous assemblies too incomplete, then the current stimulus would not lead to a fully organized assembly, which implies that the perceiver would never fully be able to see all parts of the stimulus.

The analogy to Hebb's theorizing is useful, not because these data provide any kind of test of his work, but that there seems to be some kind of formal similarity between the concepts of growth in perceptual experience (ontogenesis) that Hebb speculates about, and the development of a percept (microgenesis) that was actually measured in this experiment.

REFERENCES

Eriksen, C. W. Unconscious processes. In M. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.

Haber, R. N., & Hershenson, M. Effects of repeated brief exposures on the growth of a percept. *J. exp. Psychol.*, 1965, **69**, 40–46.

Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.

Hershenson, M., & Haber, R. N. The role of meaning on the perception of briefly exposed words. *Canad. J. Psychol.*, 1965, in press.

Pierce, J. Determinants of threshold for form. *Psychol. Bull.*, 1963, **60**, 391–407.

Thorndike, E. L., & Lorge, I. *A teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

Webster's new international dictionary of the English language. (2nd ed., Unabridged) Springfield, Mass.: G. & C. Merriam, 1939.

(Received January 20, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 287–291

EFFECT OF CS-ONSET UCS-TERMINATION DELAY, UCS DURATION, CS-ONSET UCS-ONSET INTERVAL, AND NUMBER OF CS-UCS PAIRINGS ON CONDITIONED FEAR RESPONSE [1]

ANDREW STROUTHES

Harpur College, State University of New York

120 rats in 2 3 × 3 factorial designs were used to study the effects of CS-onset UCS-termination delay and of the number of training trials on conditioned fear response. CS-onset UCS-termination delays of .55, 1.10, and 2.20 sec.; and 8, 16, and 32 CS-UCS pairings were used in both designs. A .25-sec. UCS duration was constant in Design I, whereas in Design II a .30-sec. CS-onset UCS-onset interval was constant. The effects were measured in terms of running speed to the goal box in a well-learned runway into which the CS, light previously paired with shock, was introduced as obstruction. Fear strength and CS-onset UCS-termination length were inversely related, a finding interpreted as not supporting the 2-process theory. The other factors were not significant.

It has been found that delay of reinforcement, among other factors, influences the strength of a conditioned motor response (CMR). The relationship is inverse, and the assumed mechanism has been drive-stimulus reduction (Hull, 1943). Some advocates of this explanation maintain that drive-stimulus reduction is also responsible for conditioning of fear responses. On the other hand, Mowrer (1950) posited drive-stimulus reduction for CMRs and contiguity for conditioned fear responses (CFRs).

Whether drive-stimulus reduction affects learning of fear responses might perhaps be studied by manipulating CS-onset UCS-termination delays (shock-termination delays, or STD). If in fear conditioning, shorter STDs bring about stronger CFRs, one might then consider the effect analogous to that of delay in reinforcement in the case of CMRs. The supposition might then be that the same mechanism is involved in the learning of both kinds of responses. Whereas some writers (Runquist & Spence, 1958; Strouthes & Hamilton, 1959) found CFR and STD to be inversely related, others (Mowrer & Aiken, 1954) did not.

Conducting an experiment in which STD is varied necessarily involves two other variables, i.e., CS-onset UCS-onset interval, (ISI), and UCS duration. For, by virtue of the physical arrangements, when STD is manipulated, either of these two variables can be made to vary systematically while the other remains constant.

In addition, the question of the effect of the number of CS-UCS paired presentations on CFR has not yet been satisfactorily answered. Some investigators (Gwinn, 1951; Kalish, 1954; Libby, 1951; Strouthes & Hamilton, 1959) have found a positive relation between CFR and conditioning trials, others (Mathers, 1957) have not, and still others (Goldstein, 1960;

[1] Based on a dissertation submitted to Temple University Graduate School in partial fulfillment of the requirements for the PhD degree. The author is grateful to H. C. Hamilton for his generous assistance. A paper based on this investigation was read at the 1960 APA meeting, Chicago.

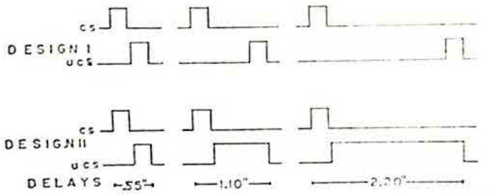


FIG. 1. A schematic representation of CS and UCS arrangements in Designs I and II. (The three levels of CS-onset UCS-termination delay in the two designs are the same—.55, 1.10, 2.20 sec. In Design I, the UCS duration is constant, .25 sec., while the CS-onset UCS-onset interval is invariant, .05 sec., and the UCS duration necessarily changes.)

Libby, 1951) have showed a curvilinear relation. In the present investigation, this factor was also studied. It was made to constitute the vertical dimension in the two designs; it is not shown in Fig. 1.

Thus the problems under investigation were: (*a*) To find the effect of CS-onset UCS-termination interval, and (*b*) of the number of CS-UCS paired presentations on CFR. (*c*) To make some inferences regarding the effects of UCS duration, and ISI on CFR.

Method

Subjects and Design

The *S*s in the two 3 × 3 factorial designs were 120 Wistar male albino 90-day old rats. They were randomly assigned to 15 subgroups. (Since 3 of the 9 subgroups in Design I were identical with 3 of the 9 subgroups in Design II, only 15 subgroups of eight rats each were used.) The horizontal dimension in Fig. 1 represents the three STD variations which are referred to as Groups AI, BI, CI for Design I and AII, BII, and CII for Design II.

In both designs the STD values were .55, 1.10, and 2.20 sec. In Design I a .25-sec. UCS duration was held invariant, whereas the ISI varied. Thus, the ISIs were necessarily .30, .85, and 1.95 sec. In Design II the ISI was held constant at .30 sec., and the UCS duration necessarily became .25, .80, and 1.90 sec. The CS-UCS paired presentations in both designs were 8, 16, and 32.

Apparatus

Runway.—A straight 48 × 4 × 4 in. runway with attached 7 × 7 × 4 in. start and goal boxes was used. The boxes were separated from the runway proper by guillotine doors. A 7-w. bulb was mounted within the runway proper, 16½ in. from the food cup or end wall of the goal box allowing 3¼ in. clearance from the floor of the runway. The light, off during training trials, could be made to blink at 16 cps during the test trials. Time was measured by a chronoscope controlled by the weight of the animal on two plates flush with the floor of the runway. They were located just outside the start box and in the goal box, 47 in. apart. The runway, the start, and goal box were painted flat white, flat gray, and flat black, respectively.

Shock box.—The *S*s were shocked in a 17 × 5 × 7 in. wooden box. A 16-cps blinking light from a 7-w. bulb mounted so that its bottom touched a Plexiglas ceiling 3¼ in. above the grid served as CS during fear conditioning. A Stoelting automatic timer controlled the lengths of the CS, UCS, and the interval between them. The UCS, 290 μa. electric shock, was supplied from an automatic compensator. It consisted of interrupted direct current with on and off durations of .014 and .041 sec., respectively.

Procedure

Runway training.—On the first day at 48 hr. food deprivation, each rat was put in the start box for 20 sec. The *S* was then lifted out and placed in the goal box whence it was removed when the three 4 × 3.3 mm., 45-mg. Noyes tablets were consumed. This procedure was repeated two more times after which the running trials began. There was a total of 37 such trials: 2 on Day 1, 5 on Day 2, and 10 on each of 3 subsequent days. Twenty minutes after the final daily trial, *S* was allowed a 30-min. feeding. The *S*s were run in virtual darkness.

Fear conditioning.—For CS-UCS relations see Fig. 1. Concerning CS-UCS pairings, since the groups received varying numbers, a procedure was adopted which allowed equal handling of *S*s and constant time interval between the last fear conditioning trial and testing. Thus, *S*s in the eight pairings groups, were fear conditioned on the last 2 fear conditioning days, whereas *S*s assigned to 16 were conditioned on the last 4 days. On days before fear was conditioned each *S* was simply placed in the shock apparatus for 5 min. without light or shock.

Testing.—The operations during testing were the same as during runway training except for the presence of the CS, which began blinking as the start-box door was raised, and went off when the animal reached

the goal. A 2-min. criterion and a 1-min. intertrial interval were adopted. The *S*s were tested on 3 consecutive days at 10, 10, and 5 trials daily. Differences in running speeds were taken as an index of the differential CFR strength acquired during fear conditioning.

Results

Design I.—All data were transformed into running speeds, $\frac{1}{\text{sec.}} \times 100{,}000$. Comparisons of the mean speeds based on the last four runway training trials for the nine subgroups resulted in nonsignificant *F*s. A similar analysis of the first test-trial data showed that mean running speeds differed significantly, $F(2, 63) = 3.30$, $p < .05$, with respect to STD only. As shown in Table 1, the differences between Groups AI and BI, and AI and CI were significant at $p = .05$ and $p < .05$, respectively. There was no difference between BI and CI. This effect is shown in Fig. 2 where *S*s from the two longer STD groups, BI and CI, reach the goal box faster than those from AI. There were no differences, with regard to this variable beyond the first test trial. The groups did not differ according to CS-UCS pairings on any trial.

Design II.—The statistical procedure used to analyze the data in Design II was similar to that em-

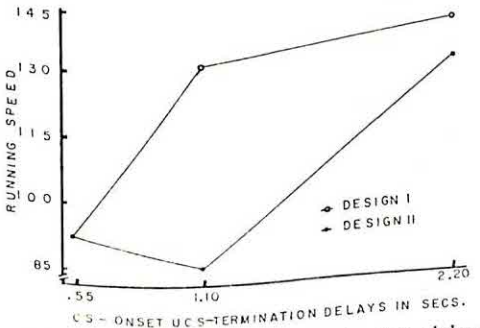


FIG. 2. CS-onset UCS-termination delay and speed of running on Test Trial 1 for Designs I and II.

TABLE 1

MEAN SPEED DIFFERENCES WITH RESPECT TO CS-ONSET UCS-TERMINATION DELAYS WITH INTRA- AND INTERDESIGN GROUP COMPARISONS ACCORDING TO THE TUKEY GAP TEST

Groups		Mean Speed Values by Groups Appearing in Column 1 and Presented in That Order		D	p
Design I	Design I				
AI	BI	93.54	133.29	39.75	.05
AI	CI	93.54	140.95	47.41	<.05
BI	CI	133.29	140.95	7.66	—
Design II	Design II				
AII	BII	93.54	86.25	7.29	—
AII	CII	93.54	132.87	39.33	.05
BII	CII	86.25	132.87	46.62	<.05
Design I	Design II				
AI	BII	93.54	86.25	7.29	—
AI	CII	93.54	132.87	39.33	.05
BI	AII	133.29	93.54	39.75	.05
BI	BII	133.29	86.25	47.04	<.05
BI	CII	133.29	132.87	.42	—
CI	AII	140.95	93.54	47.41	<.05
CI	BII	140.95	86.25	54.70	.01
CI	CII	140.95	132.87	8.08	—

ployed in Design I. As in Design I, (*a*) the groups were not reliably different at the termination of runway training. (*b*) Mean speeds for Test Trial 1 showed that STD was the only significant source of variation, $F(2, 63) = 3.23$, $p < .05$. As shown in Table 1, there was no difference between AII and BII, but there were differences between AII and CII and between BII and CII at $p = .05$ and $p < .05$, respectively. According to Fig. 2, the longest STD group (CII) ran the distance faster than either AII or BII. As in Design I, no other significant differences were found.

The identical conditions under which the two designs were conducted made interdesign comparisons possible. Thus, the mean speed differences for STD groups on Trial 1 were tested. These differences and the p values appear in Table 1. Group comparisons across designs show that

(*a*) the groups with the shortest STD in either design were significantly different from those with the longest, and (*b*) the combination of intermediate STD (1.10 sec.) and intermediate shock duration (.80 sec.) Group BII, led to slowest running speeds. This group differs at $p < .05$ from its corresponding Group BI in Design I, and at $p = .01$ from the longest STD group, CI, also in Design I. Furthermore, Groups AI and CII differ at $p = .05$, and Groups CI and AII at $p < .05$.

Discussion

CS-onset UCS-termination delay, UCS duration, ISI.—While increments in STD were identical in both designs there were also concurrent increments of ISI in Design I and of UCS duration in Design II. Therefore, no unequivocal inferences could be drawn from either design separately. However, across design group comparisons made some conclusions more definite. Thus, overall comparisons show that where given STD groups differed reliably the direction of these differences could, generally, be considered the same in both designs. Thus, running speeds increased with longer STD, longer ISI in Design I and with longer STD, longer UCS duration in Design II. If ISI is assumed to play the important role, then the outcome in Design I was in the expected direction. However, in Design II even though the ISI was constant, the groups still differed. Alternatively if we assume UCS duration to play the important role in fear conditioning, then in Design II CFR strength might be expected to increase as a function of UCS duration. Yet this was not the case. Furthermore, there was no difference between CI and CII though *S*s in the latter group were given almost eight times as long a shock as those of the former. The difference between BI and BII might be attributed either to a stronger CFR due to longer UCS in Design II or a weaker CFR due to longer ISI in Design I. The second possibility seems unlikely since in CII for instance, the ISI was even shorter than that of BI yet the CFR was weaker. Thus, whatever the effects of UCS duration and of ISI may have been they were weak at the third delay level and seem to have been offset by stronger effects of a "more important" factor, CS-onset UCS-termination delay.

In Design I, both BI and CI differed from AI. If ISI was responsible for the effects then the three groups in Design II should not differ, since the ISI was constant. Similarly, in Design II, both AII and BII differed from CII which had the longest UCS duration. In both instances these results may mean that STD is more important in fear conditioning than ISI or shock duration. BII seems to deviate from this general formulation. This finding cannot be explained readily. At best, perhaps the particular UCS duration and ISI as well as the STD for that group make for "optimal" conditions.

The STD gradient, as evidenced by the difference in running speed between the shorter and the longer STD groups, shows a steep decline. It is in agreement with the theoretical postulations by Hull (1952, p. 131).

Thus the following generalization seems parsimonious: (*a*) The delay between CS-on and UCS-off seems to play the most important role in the present experiment. (*b*) Short STDs lead to stronger CFR; increasing these, leads to rapid CFR diminution, indicating a steep gradient. (*c*) When STD is "long" the effect of UCS duration is more than offset by the effect of STD. However, (*d*) a short STD and a relatively "long" shock, as in BII, bring about strongest CFR, thus making UCS duration effective only when it combines with relatively short STD. (*e*) ISI does not seem to be as effective as STD or UCS duration in the present study. STD is the interval separating drive reduction from the stimulus which by arrangement of conditions comes to acquire the UCS properties. The findings support those by Sullivan (1950) and Runquist and Spence (1958). Thus, it may be concluded that

drive-stimulus reduction affects learning of responses, be they motor or fear, and so perhaps no other principle is needed.

Number of CS-UCS paired presentations.—As noted earlier, the relation between CS-UCS paired presentations on CFR has not yet been established. In this study no significant group differences have been obtained. It is possible that with 40 reinforced runway training trials the strength of the approach response masked any differences due to CFR. It is also possible that, in part at least, the CS duration may have been responsible. For, whenever the CS went on during fear training, it lasted .25 sec. only, whereas on any test trial the CS, acting as obstruction, lasted a considerably longer time, thus enhancing CFR extinction which in turn "erased" any group differences due to pairings.

Though there were no significant group differences there was a strong suggestion that as CS-UCS pairings increased beyond a certain number the CFR decreased. Current work supports the desensitization explanation.

REFERENCES

GOLDSTEIN, M. L. Acquired drive strength as a joint function of shock intensity and number of acquisition trials. *J. exp. Psychol.*, 1960, **60**, 349–358.

GWINN, G. T. Resistance to extinction of learned fear drives. *J. exp. Psychol.*, 1951, **42**, 6–12.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KALISH, H. I. Strength of fear as a function of the number of acquisition and extinction trials. *J. exp. Psychol.*, 1954, **47**, 1–9.

LIBBY, A. Two variables in the acquisition of depressant properties by a stimulus. *J. exp. Psychol.*, 1951, **42**, 100–107.

MATHERS, B. L. The effect of certain parameters and the acquisition of fear. *J. comp. physiol. Psychol.*, 1957, **50**, 329–333.

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.

MOWRER, O. H., & AIKEN, E. G. Contiguity vs. drive-reduction in conditioned fear: Variations in conditioned and unconditioned stimuli. *Amer. J. Psychol.*, 1954, **67**, 26–38.

RUNQUIST, W. N., & SPENCE, K. W. Temporal effects of conditioned fear on the eyelid reflex. *J. exp. Psychol.*, 1958, **55**, 613–616.

STROUTHES, A., & HAMILTON, H. C. Fear conditioning as a function of the number and timing of reinforcements. *J. Psychol.*, 1959, **48**, 131–139.

SULLIVAN, J. J. Some factors affecting the conditioning of the galvanic skin response. Unpublished doctoral dissertation, State University of Iowa, 1950.

(Received January 22, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 292–297

EFFECTS OF DRIVE, REINFORCEMENT SCHEDULE, AND CHANGE OF SCHEDULE ON PERFORMANCE [1]

PIETRO BADIA

Bowling Green State University

Acquisition of a runway response by 71 albino rats in different segments of a straight alley was investigated as a function of percentage of reinforcement (50% or 100%), 3 drive levels, and shifts of reinforcement schedule. Preshift asymptotic analyses revealed a significant drive effect but no reinforcement schedule or Drive × Reinforcement interaction effect. Postshift trials (101–120) within group comparisons indicated that high- and medium-drive, continuously reinforced Ss shifted to partial increased speed of responding on 2 response measures but under low drive no change on any measure occurred. No change occurred for high- and medium-drive, partially reinforced Ss shifted to continuous reinforcement, but for low-drive Ss an increase in running speed occurred. Results are discussed in terms of Spence and Amsel's recent theoretical statements.

The present study was designed to investigate the effects of drive level and partial reinforcement on behavior during the acquisition of a running response. A second purpose was to determine the effects of shifting from partial to continuous reinforcement in some animals and from continuous to partial reinforcement in others.

Several studies (Goodrich, 1959; Wagner, 1961) have provided support for the theoretical prediction of Amsel (1958) and Spence (1960) that partially reinforced *S*s will perform at higher levels than continuously reinforced *S*s. The basis for this predicted superiority is that nonreinforcement and the subsequent development of conditioned frustration (r_f) results in an increased motivational level. It is not known, however, how drive affects the development of r_f. Apparently, only two studies have been reported dealing with partial reinforcement and drive level (Lewis & Cotton, 1957; Linton & Miller, 1951). Drive was manipulated only during extinction in the Linton and Miller study; in the Lewis and Cotton study training was terminated before asymptotic measures were obtained and instead of response segments, total running time was recorded.

METHOD

Subjects.—The *S*s were 71 male albino rats of Sprague-Dawley strain ranging in age from 90 to 120 days at the start of training.

Apparatus.—The apparatus was a straight enclosed plastic runway 30 in. long from start box to goal box. A more complete description is given elsewhere (Ehrenfreund & Badia, 1962). Both start and goal boxes were 15 in. long and all sections of the runway were 2.5 in. wide and 4 in. high. The sides, bottom, and top were constructed of .125-in. plastic. Guillotine-type translucent doors 15 in. from each end of the runway created a start box and goal box. In front of each of these doors was another horizontally sliding door of .125-in. clear plastic. The walls, floor, and ceiling of the runway (except for slits for the doors) were continuous, homogeneous, and unbroken from one end of the runway to the other. To avoid the animal slipping, the

[1] This paper is based upon a dissertation submitted to the Graduate School of Adelphi College in partial fulfillment of the requirements for the PhD degree. The author wishes to thank David Ehrenfreund under whose direction this investigation was conducted. The research was performed while the author held a predoctoral fellowship No. MF 12,457 from the National Institute of Mental Health.

plastic floor was scored from end to end with .25-in. squares and sandblasted.

Filtered (Wratten 25A) light beams were directed from above the maze to photoelectric cells below the maze. Both the light sources and photocells were on tracks. The apparatus was located in a sound-deadened room illuminated by one $7\frac{1}{2}$-w. bulb with a random noise generator providing white noise for masking purposes. The maze was illuminated by four $7\frac{1}{2}$-w. bulbs along each side with the illumination level maintained at 40% by a variac. Starting, running, and goal measures were provided by three timing circuits activated by the interruption of the filtered photocell light sources. The first running time was measured from the interruption of the first beam 4 in. from the start door to the interruption of the second beam 16 in. from the start door and 12 in. from the first beam. The second running-time measure was taken from the interruption of the second beam to the interruption of the third beam 28 in. from the start door and 12 in. from the second beam. Goal-box time was measured from the interruption of the third beam to interruption of the fourth beam 40 in. from the start box and 12 in. from the third beam. This last beam was 7 in. from the end of the runway and 6 in. from the food cup.

Design.—Three levels of drive and two percentages of reward were combined in a 2 × 3 factorial design. The drive levels employed were $2\frac{1}{2}$ hr., $11\frac{1}{2}$ hr., and $22\frac{1}{2}$ hr., of deprivation. In addition the high-drive *S*s ($22\frac{1}{2}$ hr.) were kept at 85% of their ad-lib. body weight. The percentages of reinforcement employed at each level of drive were 100 and 50%. Twelve *S*s were randomly assigned to each of the six cells of the 2 × 3 design. However, due to sickness (1 *S*) or failure to run (4 *S*s) 5 *S*s were discarded, 4 of which were replaced by reserve animals on hand for this purpose, resulting in 12 *S*s in each cell except for the high-drive ($22\frac{1}{2}$ hr.), 100% reinforcement cell which had 11 *S*s.

Pretraining.—Fifteen days before the first acquisition trial all *S*s were placed on a $22\frac{1}{2}$-hr. food deprivation schedule and continued on it throughout the course of the experiment.

Five days prior to experimental training each *S* was introduced to 45-mg. pellets (Noyes) by receiving 30 of these pellets immediately before his regularly scheduled meal. On Day 2 this ration was reduced to 20 pellets, on Day 3 to 10 pellets, and 5 pellets were then given on the remaining 2 days. Two days prior to the first acquisition trial each *S* was allowed 5 min. to explore the runway.

Experimental preshift training.—The *S*s were tested in squads of six, one *S* from each of the experimental groups. The *S*s in the 100% groups were reinforced on all trials with four pellets (Noyes). The *S*s in the partially reinforced groups (50%) also received four pellets each in the following sequence of reinforced and nonreinforced trials: + − + −, − + − +, + − − +, − + + −. On Days 1 and 2 only one trial a day was given to each *S*, on Day 3 two trials were given, and beginning on Day 4 and thereafter all *S*s were given four trials a day.

After *S* was placed into the starting box, during a training trial, the vertical door was raised when *S* faced it and 2 sec. later, the horizontally sliding clear plastic door was opened. When *S* left the start box, the vertical door was lowered to prevent retracing. When *S* entered the goal box, the vertical door was lowered and after eating the food, removed. On nonreinforced trials *S* was removed after 10 sec. After all *S*s had received 1 trial the sequence was repeated until each had 4 trials. About 45 min. after the experimental session was completed, *S*s were returned to their home cage. A total of 88 acquisition trials was given each *S*.

Experimental postshift training.—After the 88 acquisition trials the percentage of reinforcement was reversed for all *S*s. The 100% reinforced groups were reduced to a 50% reward schedule and the 50% reinforced groups shifted to a 100% reward schedule. An additional 32 trials were then run under the conditions that prevailed prior to the shift.

Results

The running and goal speeds were transformed to reciprocals and the performance measures were separately analyzed for the three phases of training: initial preshift trials (1–20); terminal preshift trials (69–88); postshift trials (101–120).

Preshift trials.—From Tables 1, 2, and 3 it appears that in all three preshift response measures during the early trials (1–20), continuous reinforcement produced faster running speeds than partial reinforcement while high drive produced faster running speeds than medium drive, which, in turn, produced faster running speeds than low drive. For all

TABLE 1

MEAN RUNNING SPEED IN THE FIRST SECTION OF THE ALLEY AS A FUNCTION OF DRIVE, PERCENTAGE OF REINFORCEMENT, AND TRIALS

Drive Level	Percent Reward	Trials 1–20	Trials 69–88	Trials 101–120[a]
High	100	1.767	3.335	3.358
	50	1.402	3.493	3.486
Medium	100	1.462	2.870	3.125
	50	1.239	3.169	3.179
Low	100	1.275	2.307	2.390
	50	1.033	2.247	2.365

[a] Postshift trials.

three response measures the higher drive groups appeared to have maintained their superiority over the lower drive groups in late acquisition training also. However, at the terminal phase of training for all drive levels, the marked superiority of continuously reinforced groups over the partially reinforced groups was considerably reduced for the goal-response measure, eliminated for the second speed measure, and apparently reversed for the high- and medium-drive groups in the first section of the runway (first speed measure).

Mean running speeds at three levels

TABLE 2

MEAN RUNNING SPEED IN THE SECOND SECTION OF THE ALLEY AS A FUNCTION OF DRIVE, PERCENTAGE OF REINFORCEMENT, AND TRIALS

Drive Level	Percent Reward	Trials 1–20	Trials 69–88	Trials 101–120[a]
High	100	1.900	3.271	3.480
	50	1.522	3.372	3.344
Medium	100	1.605	2.859	3.052
	50	1.402	3.101	3.085
Low	100	1.374	2.315	2.377
	50	1.161	2.254	2.366

[a] Postshift trials.

of drive and two percentages of reinforcement were separately analyzed for the three response measures over the initial trials (1–20) and terminal trials (69–88). The separate analyses for the initial trials showed that: (*a*) percentage of reinforcement was significant for the first, second, and goal-speed measures, $F\ (1, 65) = 32.60$, 28.06, and 29.51, respectively, $p < .005$; (*b*) drive level was significant for all three response measures, $F\ (1, 65) = 26.36$, 26.25, 18.66, $p < .005$; (*c*) the interaction between

TABLE 3

MEAN RUNNING SPEED IN THE GOAL SECTION OF THE ALLEY AS A FUNCTION OF DRIVE, PERCENTAGE OF REINFORCEMENT, AND TRIALS

Drive Level	Percent Reward	Trials 1–20	Trials 69–88	Trials 101–120[a]
High	100	1.642	3.151	3.261
	50	1.286	3.004	3.074
Medium	100	1.424	2.715	2.724
	50	1.143	2.813	2.879
Low	100	1.177	2.186	2.214
	50	1.046	2.039	2.238

[a] Postshift trials.

percentage of reinforcement and drive level was not significant on any of the three response measures.

Terminal trials.—The analyses for the terminal trials of acquisition training showed that (*a*) percentage of reinforcement did not attain significance on any of the response measures; (*b*) drive level was significant for the first, second, and goal measures, $F\ (1, 65) = 34.07$, 34.08, 42.84, respectively, $p < .005$; (*c*) in none of the three response measures was there a significant percentage of reinforcement and drive level interaction.

As seen from Tables 1, 2, and 3, the net effect of shifting the percentage

of reinforcement was a convergence of the postshift running speeds for the continuously and partially reinforced *S*s. Since terminal preshift asymptotic performance levels between continuously and partially reinforced *S*s did not differ significantly, it is obvious that the converging postshift running speeds would not differ. However, differential within-group effects resulting from shifting the percentage of reinforcement did occur. Under conditions of high- and medium-drive levels, continuously reinforced *S*s shifted to a partial schedule, increased their speed of responding on the first and second running speed measures, but not on the goal-speed measure. However, under conditions of low drive, little change resulted from the shift on any of the response measures. It also appears from Tables 1, 2, and 3 that under conditions of high and medium drive there is little effect of shifting partially reinforced *S*s to a continuous schedule. But under conditions of low drive, shifting from a partial to continuous schedule resulted in an increase in running speed on all three response measures.

Since separate statistical analyses involving percentage of reinforcement, drive level, and pre- and postshift trial blocks (last 20 pre- and postshift trials), resulted in a significant triple interaction for the first and second speed measure, separate analysis of the preceding observations was computed for all three response measures.

The results of these within analyses of partially reinforced *S*s shifted to continuous reinforcement showed no significant differences between pre- and postshift speed measures under conditions of high and medium drive. Under conditions of low drive partially reinforced *S*s, shifted to a continuous schedule, did increase their speed of responding on all three response measures, $F\ (1, 11) = 11.05$, $p < .025$; $F\ (1, 11) = 19.38$, $p < .005$; and $F\ (1, 11) = 31.72$, $p < .005$ for the first, second, and goal-speed measures, respectively.

The analyses of the high-drive, continuously reinforced *S*s shifted to a partial schedule showed that: (*a*) pre- and postshift running speed differences for the first response measure were significant, $F\ (1, 10) = 12.14$, $p < .005$, reached significance in the second response measure, $F\ (1, 10) = 6.32$, $p = .025$, but did not attain significance for the goal-response measure, $F\ (1, 10) = 3.35$, $p < .10$; (*b*) under conditions of medium drive also the postshift running speeds were superior to preshift speeds for the first and second response measures yielding $F\ (1, 11) = 14.15$, $p < .005$ and $F\ (1, 11) = 12.83$, $p < .005$, respectively, but did not approach significance for the goal-response measure ($F < 1$); (*c*) with none of the three response measures did low-drive *S*s show a significant change in speed of responding when shifted from continuous to partial reinforcement.

Discussion

Several of the findings in this study were expected. The second-order interaction between percentage of reinforcement and trials predicted by the theories of Amsel (1958) and Spence (1960), reported by other investigators (Goodrich, 1959; Wagner, 1961), and obtained in this study was significant in both the pre- and postshift phases of training. This interaction indicates that percentage of reinforcement affected the rate of acquisition training. But in none of the preshift performance measures was there a significant triple interaction. Failure to obtain this interaction suggests that the differential effects of percentage of reinforcement on performance are independent of drive level. It should be noted in regard to this latter finding that

there is nothing in the theories cited (Amsel, 1958; Spence, 1960) to suggest findings contrary to this.

The acquisition running speeds for the first and second measures follow a pattern similar to that reported by others (Goodrich, 1959; Haggard, 1959; Wagner, 1961). Although partially reinforced groups performed at an inferior level early in training, it is clear that following extended training they attain a level of performance if not greater, at least equal to, continuously reinforced *S*s. Despite the fact that the apparent reversal of initial order for the high- and medium-drive groups was not significant, the direction is consistent with that of other studies where statistical support was obtained (Goodrich, 1959; Wagner, 1961). These results provide some support for the theoretical formulation of Amsel (1958) and Spence (1960).

Contrary to the preshift findings, the analysis using pre- and postshift trial blocks indicated a lack of independence between percentage of reinforcement and drive level. While the shift from continuous to partial reinforcement resulted in an increase in running speed for high- and medium-drive groups on the first two response measures, no increase occurred for the low-drive group. In contrast to these findings, the shift from partial to continuous reinforcement led to a significant increase in running speed for the low-drive group, but not for high- and medium-drive groups. It should be noted that the earlier preshift relationship between percentage of reinforcement and drive level, though lacking statistical support, was in the same direction as the latter findings.

The findings of the postshift analysis suggest two possible extensions of the theoretical statements of Amsel (1958) and Spence (1960) regarding the conditioned form (r_f) of the hypothetical emotional response termed *frustration*. In reference to the first extension, the findings of the present investigation imply that under conditions of low drive and small reward r_f may develop only minimally or not at all. Only under intermediate or high drive may r_f develop sufficiently to affect behavior. At present the theoretical statements offered by Amsel (1958) and Spence (1960) cannot adequately explain these data in that they assume r_f, resulting from nonreinforcement, simply to be a function of the vigor or strength of r_g. The development of r_g, in turn, is considered to be independent of drive and dependent only upon the characteristics of the reinforcement employed, such as magnitude and delay of reward. Since partially reinforced low-drive *S*s when shifted to continuous reinforcement showed a significant increase in running speed, while those shifted from continuous to partial reinforcement did not, extension of these theoretical statements to include the drive dimension in regard to the development of r_f or r_g may be necessary.

The second extension suggested by the findings deals with the extinction of the r_f response. Since the acquisition of r_f is based upon the laws of classical conditioning, presumably these laws should also apply to the extinction of r_f. Since withholding the unconditioned stimulus leads to the extinction of the conditioned response, eliminating the unconditioned stimulus (in this case, nonreinforcement) should result in the extinction of r_f and loss of the motivational component contributed by it. A decrement in response strength following the extinction of r_f should reflect this reduced motivational level.

The results of this investigation are not in agreement with the above theoretical outline. While *S*s shifted from continuous to partial reinforcement (32 trials) showed the predicted increment in running speed, *S*s shifted from partial to continuous reinforcement (32 trials) failed to show a decrement in running speed. It may be that while 32 trials are sufficient to acquire the r_f response, considerably more trials may be necessary to extinguish this emotionally based response.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

EHRENFREUND, D., & BADIA, P. Response strength as a function of drive level and pre- and postshift incentive magnitude. *J. exp. Psychol.*, 1962, **63**, 468–471.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

HAGGARD, D. F. Acquisition of a simple running response as a function of partial and continuous schedules of reinforcement. *Psychol. Rec.*, 1959, **9**, 11–18.

LEWIS, D. J., & COTTON, J. W. Learning and performance as a function of drive strength during acquisition and extinction. *J. comp. physiol. Psychol.*, 1957, **50**, 189–194.

LINTON, H. B., & MILLER, N. E. The effect of partial reinforcement on behavior during satiation. *J. comp. physiol. Psychol.*, 1951, **44**, 142–148.

SPENCE, K. W. *Behavior theory and learning: Selected papers.* Englewood Cliffs, N.J.: Prentice-Hall, 1960.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

(Received January 21, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 298–303

PERCEPTION OF OFF-SIZE VERSIONS OF A FAMILIAR OBJECT UNDER CONDITIONS OF RICH INFORMATION[1]

SAMUEL FILLENBAUM,[2] H. RICHARD SCHIFFMAN, AND JAMES BUTCHER[3]

University of North Carolina

An experiment was conducted replicating and extending an earlier study by Slack in which size estimates of off-size versions of a familiar object had been found to regress to familiar size, under conditions of rich environmental information. Size and distance estimates were obtained for normal, oversized, and undersized chairs (and control stakes and abstract constructions) located on a dirt road at various distances from S. There was no tendency for either size or distance judgments to be systematically biased as a function of familiar size; Slack's findings were not replicated. These results do not support any position which argues that familiar size is one of the major operative determinants of apparent size, under ordinary circumstances of observation in a richly informative environment.

Two rather different approaches to the determinants of apparent size and distance may be distinguished. On the one hand there is the approach to be found in Gibson (1951) which stresses the rich psychophysical information usually present and minimizes the role of known, familiar object size; on the other hand there is the transactionalist approach (see e.g., Ittelson, 1962) which emphasizes the role of past experience in building up in *O* an object representation, so that the object's dimensions are known once it is recognized. Each approach has led to experimentation which shows that the sort of variables it considers may provide *sufficient* conditions for size and distance perception. Thus, for example, Gibson (1951) has done work that indicates that the size of anonymous objects, such as stakes, may be accurately judged, even at great distances, if they are placed in the richly informative environment of a plowed field. The transactionalists have shown that under conditions where all environmental information is eliminated and only object information remains (e.g., one-eyed viewing in a completely dark surround) *O* will rely upon his past experience with an object in determining its apparent size and distance (Ittelson, 1951). But what happens when both kinds of information are available, when there is both rich information from the environment in which the object is located and the object is a familiar well known one, with matters so arranged, by a systematic distortion of object size, that these sources of information are in conflict? Using chairs of different sizes in an open field situation Slack (1956) found regressions to familiar size which were statistically significant, though rather small in absolute size. Relative to a control stake, an oversized chair was underestimated and an undersized chair was overestimated.

Characteristically, familiar objects are viewed under conditions of rich environmental information so the

[1] The authors are greatly indebted to R. D. Bock and S. Zyzanski for aid in the analysis of the data on the UNC computer.

[2] Visitor at the Center for Cognitive Studies, Harvard University.

[3] Now at the University of Minnesota.

question of the determinants of perceived size under such circumstances is an important one. There is little direct evidence on the issue and therefore this study was designed as a replication and extension of Slack's experiment. In addition to a number of procedural changes, such as complete randomization in the order of presentation of objects, this study extends Slack's experiment in two main respects: (*a*) Additional control objects were used. Slack used stakes as his control objects, but a stake differs from a chair not only in its lack of characteristic size but also in being visually less complex. A stake consists of one vertical only; whereas a chair has a number of vertical and horizontal sides. In this study, in addition to stakes, abstract constructions with a number of verticals and horizontals were used as control objects with the aim of more closely approximating chairs in regard to visual complexity. (*b*) Distance estimates were obtained concurrent with the size judgments. This was done in order to determine whether there was any systematic bias in the apparent distance of the off-size familiar objects, and to permit an examination of the relation between apparent size and apparent distance in terms of the size-distance invariance hypothesis (Kilpatrick & Ittelson, 1953). For example, if an undersized chair is seen as larger than a control object (of the same size at the same distance) will it also be seen as farther away?

Method

Subjects.—The *S*s were met at a university building, were blindfolded, and driven to the site of the experiment. There were 19 *S*s, volunteer male college students from the introductory course in psychology. The data from 3 additional *S*s were discarded since their distance estimates were quite erratic, thus an object at 30 yd. might be estimated as closer than one at 20 yd. or farther away than one at 40 yd.

FIG. 1. Medium size abstract construction at site of experiment.

Apparatus.—Three wooden chairs, three wooden stakes, and three abstract constructions constituted the experimental materials. The chairs were 56½, 34, and 25 in. in height, and were proportional in length and width. As a result of some accidents in carpentry our objects were slightly smaller than those employed by Slack which were 58, 34½, and 25½ in. in height, respectively. As in the Slack study, the chairs were armless and had two horizontal slats in the back. The stakes were 2 × 4 in. pieces of wood equal in height to the chairs. The abstract constructions, also equal in height to the chairs, consisted of stakes with horizontal and vertical arms so arranged as to be similar to the chairs in visual complexity i.e., in the number and sizes of the horizontals and verticals. A photograph of one of the abstract constructions in situ is shown in Fig. 1. All objects were painted a dark brown and were stored some distance behind *S* between trials, *S* being firmly instructed never to look behind him. An adjustable cloth tape attached to the ground at one end (and painted brown on *S*'s side) was used to obtain the size estimates.

The experimental site was an unused, closed off dirt road several hundred yards in length, and approximately 20 ft. in width; there was a wide open field and a few bushes

on one side of the road and trees on a bank on the other side. The trees were not distinctly separate or of any specific and therefore informative size; rather they blended into a quite homogeneous mass. The view was such that no artifacts such as telephone poles or houses were in sight during direct inspection of test objects. The field of view afforded rich environmental information via many gradients of texture, there was the dirt of the road with pebbles and small rocks scattered irregularly, there was the grass, and there were the bushes and trees on the side of the road.

Procedure.—The *S*s viewed all objects facing in the same direction. Objects were presented at 20, 30, and 40 yd. from *S*. Both size and distance judgments were made on each object at a given distance before it was removed and another object presented. Half of the *S*s made the size judgment for each object before the distance judgment, this order was reversed for the other *S*s. The order in which the various objects were presented at the various distances was randomly determined for each pair of *S*s, since two *S*s were run in each session, great care being taken that no communication of any sort take place between *S*s. While one *S* was giving his responses the other sat on the side, blindfolded.

Size estimates were recorded as follows: *E* stood 3 ft. in front and 5 ft. to the right of *S* and moved the comparison tape up or down, as directed by *S* by means of an upward or downward motion of the hand, until the tape appeared to *S* to have been extended to the same height as the object. This height was then recorded by *E* to the nearest half inch. In the case of distance estimates *S* was asked to estimate how far away, to the nearest yard, the object appeared to be, and to record each estimate on a fresh page of a booklet with which he had been provided.

The *S*s were told to judge the height and distance of the objects from them. They were asked to make their judgments "in the same manner as you would judge the height and distance of things every day." Essentially the instructions for height estimation were similar to those used by Slack, and instructions for distance estimation were similar to those for height estimation with appropriate modifications.

The *S*s were required to replace their blindfolds after making their judgments for each object, and remained blindfolded while one object was being replaced by another object.

Results and Discussion

The main findings are shown in Table 1 which gives the mean apparent height or size, and mean apparent distance from *S*, for each object. Analyses of variance were

TABLE 1

Mean Apparent Size and Mean Apparent Distance

Object	Distance in Yd.					
	20		30		40	
	Size in In.	Distance in Yd.	Size in In.	Distance in Yd.	Size in In.	Distance in Yd.
Large (56½ in.)						
Chair	64.79	15.37	64.32	23.16	63.79	33.89
Abstract	62.89	15.47	62.84	24.05	62.47	33.11
Stake	63.89	15.26	62.89	24.05	63.47	36.37
Medium (34 in.)						
Chair	43.21	16.37	42.84	25.74	43.00	36.05
Abstract	43.00	15.74	42.68	24.74	43.05	36.11
Stake	43.32	17.05	42.63	24.94	43.53	36.32
Small (25 in.)						
Chair	33.37	17.26	34.21	26.58	35.47	36.16
Abstract	33.84	15.63	35.21	25.63	35.21	35.11
Stake	32.47	16.42	33.58	25.68	36.00	36.74

performed on the size and distance judgments and yielded the following results.

Size judgments.—There was no significant main effect due to Distance, or due to Objects; there was a significant main effect due to *S*s, $F (18, 396) = 47.85, p < .01$; and due to Size, $F (2, 36) = 754.77, p < .01$, with larger objects being judged larger. In addition, the following interactions were also significant—*S*s $\times$ Size, $F (36, 396) = 17.20, p < .01$; *S*s $\times$ Distance, $F (36, 396) = 2.63, p < .01$; Size $\times$ Distance, $F (4, 396) = 5.96, p < .01$; and Size $\times$ Objects, $F (4, 396) = 2.95, p < .05$. With regard to the two significant interactions not involving *S*s we note, for the Size $\times$ Distance interaction, that there is a tendency for small objects to be judged somewhat larger the farther away they are; there is no such tendency for the medium or large objects. This result is somewhat different from that obtained by Slack who found an overall significant tendency for objects to be judged somewhat larger the farther away they were. But the critical interaction is that between Size and Objects, which indeed is significant ($p < .05$). An inspection of Table 1 reveals, however, that this interaction is significant *not* because oversized chairs are underestimated and undersized ones overestimated relative to their control objects, as was found by Slack, but rather because oversized chairs are somewhat overestimated relative to both of their control objects, at each of the three distances. No consistent tendency at all is apparent when judgments of medium or undersized chairs are compared with those of their control objects.

Distance estimates.—There were significant main effects due to Distance, $F (2, 36) = 174.74, p < .01$; due to Size, $F (2, 432) = 15.51, p < .01$; due to Objects, $F (2, 432) = 3.51, p < .01$; and due to *S*s, $F (18, 432) = 143.39, p < .01$. Only one of the interactions, that between *S*s and Distance, was significant, $F (36, 432) = 11.20, p < .01$. The main effect of Distance indicates merely that estimates of apparent distance varied with actual distance. The main effect due to Size reflects principally the fact that large objects were seen as somewhat closer than medium or small objects at the same distance (a sort of distance constancy compromise effect). The Object effect is due to the fact that, on the average, the abstract construction is seen as somewhat closer than the chair, which in turn is seen as somewhat closer than the stake.

An examination of Table 1 indicates that there is no consistent relation between mean apparent distance and mean apparent size for the trios of objects representing each of the nine Size $\times$ Distance combinations. The correlation between judged size and estimated distance was calculated separately for each of the 27 Size $\times$ Distance $\times$ Object combinations. The correlation coefficients obtained were moderate in value (median $r = .17$), and rather homogeneous. It should be noted that from the point of view of this study interest is not in the size-distance invariance hypothesis per se, indeed given the method used in this study where size and distance estimates are correlated for objects of given size at a given distance the evidence is equivocal (see Gogel, Wist, & Harker, 1963). Rather we are interested in whether or not the size-distance relation is systematically affected when an off-sized familiar object is involved. An inspection of the correlation coefficients indicated that, in each case, the correlation between judged size and es-

timated distance for the undersized, normal, or oversized chair was very similar to that for the appropriate control objects presented at the same distance.

It is clear that our results do not replicate those of Slack, that there is no tendency for size judgments of grossly off-size versions of familiar objects to regress to familiar or known size. There are indeed some differences in procedure and setting between this experiment and that of Slack. In particular, Slack performed his experiment in a large flat field whereas this experiment was conducted on an unused dirt road with a field on one side, and trees on a bank on the other side, the view being such that no artifacts such as telephone poles or houses were in sight during inspection of the test objects. It is hard to see how such differences in setting could account for the differences in results, and if the differences in results are attributable to differences in setting and procedure, then the generality and importance of Slack's findings, of which much is made by Ittelson (1962) is certainly called into question.

It should be pointed out that while Slack obtained statistically significant results, the effects were actually quite small in absolute terms, the difference in apparent height between chairs and stakes as a percentage of the total apparent height of the chairs varying from 4% to 10%. Thus for the oversized chair, where the effect was the greatest, the average difference between it and its control stake was less than 4 in. whereas the average judged difference between the oversized chair and the medium chair was over 17 in. Thus the effect of information from the object's environment is greatly preponderant over the effect of familiar size, in the Slack study.

Slack's experiment and this extension of it represent attempts to set up a confrontation between two different classes of possible determinants of apparent size which usually yield mutually consistent information. Insofar as very gross magnification or diminution was employed in these studies it may have had the effect of forcing S to recognize that very queer special sorts of objects were being presented for judgment, leading him to make his size and distance estimates in terms of all the information that could be extracted from the environment, with the result that there is no contamination of these estimates by familiar size in this experiment (and at the most a small, though statistically significant, regression to known size in Slack's study). It might be of considerable interest to repeat this sort of experiment with more moderate though still supraliminal over- and undersizing of objects, coming from classes homogeneous as to size. If under such circumstances one were still to obtain findings similar to those of this experiment, i.e., no regression to known familiar size, then the generality of the results obtained by the transactionalists under conditions of degraded information would be very much in doubt. In fact, even for conditions of degraded information the results of the transactionalists have been challenged by the findings of a number of studies, see e.g., Hochberg and Hochberg (1952) and Epstein (1961); for a review of the relevant literature see Epstein, Park, and Casey (1961).

Certainly our knowledge of objects built up through a history of commerce with them may be sufficient to specify their sizes, as indicated not only by the work of the transactionalists but also in studies such as that by Bolles and Bailey (1956) in which it was found that Ss' estimates of the sizes of familiar objects were no less accurate with eyes closed than under a subsequent condition of visual inspection. But such findings do *not* demonstrate that under usual circumstances of visual exposure S actually uses visual information only to recognize or identify an object. The results of this experiment indicate that under conditions of rich information Ss can and do assess the sizes and distances of grossly enlarged or diminished familiar objects without bias by familiar size.

REFERENCES

BOLLES, R. C., & BAILEY, D. E. Importance of object recognition in size constancy. *J. exp. Psychol.*, 1956, **51**, 222–225.

EPSTEIN, W. The known-size-apparent-distance hypothesis. *Amer. J. Psychol.*, 1961, **74**, 333–346.

EPSTEIN, W., PARK, J., & CASEY, A. The current status of the size-distance hypothesis. *Psychol. Bull.*, 1961, **58**, 491–514.

GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1951.

GOGEL, W. C., WIST, E. R., & HARKER, G. S. A test of the invariance of the ratio of perceived size to perceived distance. *Amer. J. Psychol.*, 1963, **76**, 537–553.

HOCHBERG, C. B., & HOCHBERG, J. E. Familiar size and the perception of depth. *J. Psychol.*, 1952, **34**, 107–114.

ITTELSON, W. H. The constancies in perceptual theory. *Psychol. Rev.*, 1951, **58**, 285–294.

ITTELSON, W. H. Perception and transactional psychology. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 4. New York: McGraw-Hill, 1962. Pp. 660–704.

KILPATRICK, F. P., & ITTELSON, W. H. The size-distance invariance hypothesis. *Psychol. Rev.*, 1953, **60**, 223–231.

SLACK, C. W. Familiar size as a cue to size in the presence of conflicting cues. *J. exp. Psychol.*, 1956, **52**, 194–198.

(Received January 23, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 304–307

SPONTANEOUS RECOVERY OF LETTER-SEQUENCE HABITS [1]

ELI SALTZ

Center for the Study of Cognitive Processes, Wayne State University

Dominant letter-sequence habits were extinguished in 69 Ss and spontaneous recovery was measured 10 min., 1 hr., 2 hr., 3 hr., or 24 hr. after extinction. A significant ($p < .01$) degree of spontaneous recovery was measured across all Ss. Recovery was asymptotic by 1 hr. after extinction.

The concept of spontaneous recovery has been used by theorists in the verbal-learning area (e.g., Gibson, 1940; Underwood & Postman, 1960). However, very little direct work has been done to demonstrate the existtence and characteristics of spontaneous recovery utilizing verbal materials. Studies such as Briggs (1954) and Rothkopf (1957) have demonstrated that in a retroactive inhibition situation the A-B list recovers in strength as a function of time following A-C (or A′-C) acquisition. Such studies are only indirect evidence concerning spontaneous recovery since they are, strictly speaking, studies of the interaction over time between two competing S-R systems. No evidence appears to exist concerning the recovery of an extinguished response which has not been superseded by the strengthening of a competing verbal response to the same stimulus.

The literature in the instrumental learning and the conditioning areas, in which the spontaneous recovery concept was originally developed, is inconclusive concerning the function relating recovery to time. Ellson (1938) measured recovery between 5 min. and 3 hr. after extinction of a bar-press response in the rat and found a continuously rising function. Lewis (1956) measured recovery between 1 and 60 min. after extinction of a running response in the rat and found that recovery had apparently not reached asymptote by 60 min. after extinction. On the other hand, Graham and Gagné (1940) found that the asymptote of a recovery of a running response was reached after only several minutes.

The function of the recovery curve poses a serious problem for writers like Underwood and Postman (1960) who suggest that at least some portion of forgetting is a product of interference arising from the recovery of responses extinguished during the learning of a new response. Should recovery of verbal responses reach asymptote shortly after extinction, only a limited portion of the forgetting curve could be accounted for by means of such a mechanism.

Because of the theoretical importance of the Underwood and Postman position, the present study was designed in the context of one of the two models outlined by these writers, the letter-sequence interference model. This model states that previous experience has developed letter-sequence habits among the various letters of the language. In learning a new response, certain of these sequences will be inappropriate and must be extinguished. With time, these extinguished responses will recover in strength and compete with the newly learned sequences, causing forgetting of the newly learned sequences.

[1] The present study was supported by National Science Foundation Grant G 21476.

In the present study, the objective was to extinguish a dominant letter-sequence habit without reinforcing a competing letter sequence, and then to observe the recovery of the original sequence as a function of time.

Method

Subjects.—The *S*s were 75 students from introductory psychology at Wayne State University. These were assigned randomly to the five recovery intervals with the limitation that 15 *S*s must be assigned to each interval. Analysis of the data revealed that, due to error by *E*, 6 *S*s had not reached criterion of extinction before cessation of extinction trials, and these *S*s were eliminated from the recovery analysis.

Procedure.—The *S*s were presented the following list of 20 letter combinations: VA, CO, KA, LO, HI, GA, FU, BE, CI, HO, FI, GO, FA, DI, JA, LA, DO, JE, DE, and RA in three different orders on successive trials. Instructions were to give the first letter that came to mind on seeing each letter pair on the memory drum. For the first five trials, *E* made no response. From Trial 6 on, *E* systematically said, "Wrong" to any response made to a preselected five items. These were the critical 5 items for which spontaneous recovery was to be measured. A different 5 items were randomly selected for each *S*. To reduce the likelihood of voluntary suppression of the dominant response to the 5 critical items, *E* attempted to obscure the situation for *S* by randomly selecting an additional 5 items on each trial and saying, "Wrong" to these also. To the responses to the remaining 10 items, *E* said, "Right." Thus *S* was systematically told his response was wrong for 5 items, and sometimes told he was right, sometimes wrong, for each of the remaining 15 items.

The first five trials, during which *E* gave *S* no feedback, were used to determine the dominant response to each of the five critical items. A response was considered dominant if it occurred on at least three of the five trials. The extinction series stopped when *S* failed to give his dominant response to all of the five critical items on two successive trials. Items were presented on a Lafayette drum at a 3-sec. rate. Very few omissions occurred.

The *S*s were tested for spontaneous recovery 10 min., 1 hr., 2 hr., 3 hr., or 24 hr. after extinction. Instructions were the same as during the previous extinction series: "Give the first letter that comes to mind upon seeing the letter pair in the drum."

Results

Extinction.—The extinction rates were extremely variable between *S*s. Number of trials to extinction (omitting the first five base-line trials) ranged from 4 to 43, with a mean of 18.52 and an *SD* of 10.84. An estimate of the number of emissions of the dominant response to a critical item was obtained by averaging the number of such emissions (again omitting the five base-line trials) over the five critical items for each *S*. These scores ranged from 1.2 to 23.4, with a mean of 7.6 and an *SD* of 5.88. Obviously, some *S*s were extremely tenacious in holding onto the dominant responses while others gave up the responses readily. These responses in almost all cases completed the initial two letters so as to constitute a familiar three-letter word.

Spontaneous recovery.—Two base lines are necessary for evaluation of spontaneous recovery. The first is degree of dominance of the critical responses prior to the onset of extinction. This was readily obtained from the first five trials. Each *S*'s dominant response was determined from each item; the number of times each dominant response occurred during the five trials was calculated and divided by 5 to give a mean number of dominant responses per trial.

The second base line required is an estimate of response strength following extinction. During extinction, *S*s were run to a criterion of two trials in which none of the dominant responses was emitted to the criterion items. Performance criteria, however, are known to overestimate the relevant effect on the criterion trials. Consequently, an assumption of complete extinction on the criterion trials is likely to give a spuriously high indication of spontaneous recovery. A

TABLE 1

Mean Responses for Preextinction Level, Extinction Level, and Spontaneous Recovery above Extinction Level

	Preextinction	Extinction $(X-3)$	Net Recovery				
			10 Min.	1 Hr.	2 Hr.	3 Hr.	24 Hr.
M	4.40	1.97	0.00	0.92	1.14	.66	1.00
SD	0.31	1.31	2.60	1.84	1.81	1.59	1.29
N	69	69	14	12	15	15	13

more satisfactory base line for extinction was calculated from the following reasoning: Call the final criterion trial X and the first criterion trial $X-1$; it follows that at least one of the critical responses was emitted on Trial $X-2$, or $X-1$ would have been the final criterion trial. However, no such mathematical constraint applies to Trial $X-3$. Consequently, performance on Trial $X-3$ was used as an estimate of response strength at extinction. It is recognized that performance on this trial is probably an overestimate of such strength since three additional extinction trials follow; however, it was felt that an overestimate would be a more conservative measure since it would make for greater difficulty in demonstrating that either extinction or spontaneous recovery had occurred.

Table 1 summarizes the data relevant to extinction and recovery. As can be seen in Table 1, there was little tendency to give any but the dominant response during the preextinction period. This tendency dropped sharply by Trial $X-3$. A simple sign test shows that 90% of the 69 *S*s dropped in tendency to give the dominant response, $z = 6.51$, $p < .01$. Similarly, over all 69 *S*s, 69% show a recovery of the dominant responses following a cessation of extinction, $z = 3.01$, $p < .01$. Clearly, then, using a conservative estimate of extinction, both extinction and spontaneous recovery can be demonstrated in the present type of situation.

Net recovery, indicated in Table 1, is the extent to which tendency to give the dominant response has increased over the $X-3$ estimate of extinction. At 10 min. after extinction there is no indication of spontaneous recovery. Spontaneous recovery appears at the 1-hr. test and does not increase between this point and 24 hr. However, analysis of variance over the five recovery periods yields an insignificant F of approximately 1.00.

Finally, it is interesting to note that neither number of overt responses prior to extinction, nor level of response at the extinction base line $(X-3)$ is related to net spontaneous recovery in any of the groups.

Discussion

The present study clearly indicates that extinction and spontaneous recovery of letter-sequence habits occurs. Recovery is relatively sizable when one considers that the extinction base line is 1.97 items and the preextinction response level is 4.40 items, producing a maximum possible recovery range of 2.43 items. Compared to this maximum, the obtained recovery at 1 hr. is 38%; combining the 1-, 2-, 3-, and 24-hr. conditions, the recovery is 30%. As has been already suggested, these are probably underestimates of recovery since the $X-3$ base line underestimates the degree of extinction.

A conservative conclusion from the present data is that spontaneous recovery of letter-sequence responses is close to its obtained maximum at 1 hr. with the curve between 1 and 24 hr. being relatively flat. Since the material used in the present study elicits dominant responses that produce familiar three-letter words, test periods of longer than 24 hr. run the danger that *S*s may practice the dominant habits during their normal extra-experimental activities; measured recovery might then be spuriously high indexes of *spontaneous* recovery.

The obtained data of the present study suggest that no recovery has occurred by 10 min. after learning. Since the actual amount of extinction produced in this study is probably somewhat lower than that measured by the $X - 3$ base line employed, it appears reasonable to guess that some recovery has occurred in the first 10 min. At any rate, it appears that the critical interval for future examination is within the first hour after extinction.

The results of the present study support the suggestion by Underwood and Postman (1960) that letter-sequence habits are subject to extinction and spontaneous recovery paralleling that obtained in instrumental learning experiments. Peak and Deese (1937) have previously demonstrated extinction and spontaneous recovery of a verbal CVC response in paired associates using a procedure that parallels the classical conditioning experiment. The question remains, given that these processes occur, what is the extent of their role in acquisition and forgetting of verbal materials?

Some caution must be exhibited in relating the present results to the Underwood and Postman (1960) theory of forgetting. The present procedures appear to have resulted in a much more tenacious tendency for *S*s to overtly emit the responses being extinguished than is usually found in verbal-learning studies (i.e., in the typical verbal-learning study, almost no intrusions from outside the list occur). While the data here reported indicate no relationship between number of responses prior to extinction and net amount of recovery, the issue cannot be considered closed; it is possible that asymptote of spontaneous recovery might be reached later if extinction were more rapid. However, with this caution in mind, the data in this experiment suggest that decrements which occur in retention of newly learned responses between 1 and 24 hr. after learning cannot be attributed to increments in spontaneous recovery of the original dominant letter-sequence habits, since the spontaneous recovery curve is asymptotic over this interval. Spontaneous recovery could, of course, be relevant to forgetting which occurs within the first hour after new learning.

REFERENCES

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

ELLSON, D. G. Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GRAHAM, C. H., & GAGNÉ, R. M. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, **26**, 251–280.

LEWIS, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial interval. *J. exp. Psychol.*, 1956, **51**, 45–53.

PEAK, H., & DEESE, L. Experimental extinction of verbal material. *J. exp. Psychol.*, 1937, **20**, 244–261.

ROTHKOPF, E. Z. A deduction from an excitation-inhibition account of retroactive inhibition. *J. exp. Psychol.*, 1957, **53**, 207–213.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

(Received January 23, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 308–311

ASSOCIATIVE AND DIFFERENTIATION VARIABLES IN ALL-OR-NONE LEARNING [1]

CLESSEN J. MARTIN [2]

Center for the Study of Cognitive Processes, Wayne State University

An assessment of the stimulus-differentiation variable within the 1-trial learning controversy was made. All 20 Ss in each group learned the same final 12 paired associates to 1 perfect trial. The C group was presented the same 12 items on all trials. The E-1 group had a different pairing of the 3 S-R terms during the 1st 2 trials from that of the remaining trials. The E-2 group, during the 1st 2 trials, was presented 3 stimuli which were then replaced by the 3 stimuli learned to criterion by the other 2 groups. Only data from the 20 Ss in each group who failed to respond correctly to any of the 3 critical items during the 1st 2 test trials were analyzed. The results showed that significantly fewer number of trials were required to learn the 3 critical items by the C group as compared to the E-1 group. This finding provides evidence against the 1-trial learning position. The importance of the stimulus-differentiation variable was shown by the fact that a significantly fewer number of trials were required to learn the 3 critical items by the E-1 group in comparison with the E-2 group.

This study had two major objectives. The first objective was to provide an alternative procedure to the Rock-type design (Rock, 1957) for the investigation of one-trial learning. The proposed advantage of the present procedure is that the final list upon which the experimental and control groups reached criterion was the same. Therefore, the matter of item difficulty became less important in a comparison of the performance of the experimental group vs. the control group. The second objective was to assess the relevancy of the stimulus-differentiation variable within the one-trial learning controversy.

[1] This article is based upon a dissertation submitted to Wayne State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author gratefully acknowledges the assistance given by Eli Saltz. The investigation was supported in part by a research grant, G-21476, granted to Eli Saltz from the National Science Foundation. A report of these results was presented at the 1963 American Psychological Association meeting.

[2] Now at Michigan State University.

The general procedure in this study consisted of presenting the control group (C) with the same 12 items on all trials. Experimental Group 1 (E-1) had a different pairing of the three critical S-R terms during the first two trials from that of the remaining trials. Experimental Group 2 (E-2), during the first two trials, was presented three stimuli which were then replaced by the three stimuli learned to criterion by the other two groups.

The validity of the one-trial position was evaluated by comparing the C and E-1 groups with respect to the ease of learning the three critical items. The one-trial position was interpreted as predicting no difference between the C and E-1 groups. If stimulus differentiation is considered to be a relevant factor, the incremental and one-trial positions would predict the E-1 group to learn the three critical items sooner than the E-2 group.

Table 1 presents a summary of the differential predictions made by the

one-trial and incremental positions with respect to the relevancy of the stimulus-differentiation variable.

Method

Subjects.—A total of 75 undergraduate students enrolled in introductory psychology at Wayne State University was tested in this investigation. Only those *Ss* who failed to respond correctly to any of the three critical pairs during the first two test trials were used in the analyses.

Materials.—Three 12-item paired-associate lists were constructed. Each of the three lists had the same nine pairs. The stimuli for the nine pairs consisted of high-meaningful (*m*) disyllabic nouns as scaled by Noble (1952). The responses were single letters and were the same for all three lists. Three stimuli in each list consisted of low *m* value disyllabic nouns paired with single letters. All groups reached criterion on List 2 which consisted of the following pairs: WAGON-S, OFFICE-Y, VILLAGE-K, HEAVEN-F, INSECT-N, *SAGROLE-Z*, KITCHEN-M, *BALAP-J*, DINNER-H, JEWEL-D, *NEGLAN-Q*, GARMENT-X. The critical items in List 2 are italicized. The three critical items in List 1 were: *NOSTAW-Z*, *ZUMAP-J*, and *MEARDON-Q*. The three critical items in List 3 were: *SAGROLE-J*, *BALAP-Q*, and *NEGLAN-Z*.

Procedure.—The S-R pairs were presented at a 1-sec. rate during the learning trials on a Stowe memory drum. After each learning trial a test trial was given. Each stimulus was typed in capital letters on a 3 × 5 in. card and during the test trials the stimuli were presented one at a time and *S* had unlimited time in which to respond. The learning and test trials were alternated until *Ss* reached a criterion of one errorless test trial. The S-R pairs were presented in three orders on the learning trials and the S terms were presented in three different orders on the test trials.

Three groups of *Ss* were tested: Control group, Experimental 1 group, and Experimental 2 group.

Control group (C).—This group learned List 2 throughout all trials.

Experimental 1 group (E-1).—This group was presented List 3 during the first two learning and test trials. Beginning with Learning Trial 3, List 2 was presented for the remaining trials. The only differences between the two lists consisted of a different pairing of the stimuli and responses among the three low-*m* pairs.

TABLE 1

DIFFERENTIAL PREDICTIONS MADE BY THE ONE-TRIAL AND INCREMENTAL POSITIONS WITH RESPECT TO THE EASE OF LEARNING OF THE THREE CRITICAL S-R ITEMS AS A FUNCTION OF THE RELEVANCY OF THE STIMULUS DIFFERENTIATION VARIABLE

Stimulus Differentiation	Theoretical Acquisition of an S-R Association	
	One-Trial	Incremental
Relevant	C = E-1 < E-2	C < E-1 < E-2
Not relevant	C = E-1 = E-2	C < E-1 = E-2

Experimental 2 group (E-2).—This group was presented List 1 during the first two learning and test trials. On Learning Trial 3, List 2 was presented for the remaining trials. For this group, new low-*m* stimuli were inserted in place of the three low-*m* stimuli presented during the first two learning and test trials.

The *Ss* were assigned alternately to the C and E-2 groups as they appeared at their scheduled time of testing. A total of 20 *Ss* who failed to respond correctly to any of the critical items during the first two test trials was tested in each of the two conditions. Subsequent testing involved testing 20 more such *Ss* in the E-1 group.

Results and Discussion

The results were based upon the 20 *Ss* in each of the three groups who failed to respond correctly to any of the three critical items during the first two test trials. Five *Ss* in each of the three groups made one or more correct responses to the three critical S-R pairs on the first two test trials.

In order to determine whether the three groups were comparable in learning ability prior to the introduction of the experimental manipulation, the Kruskal-Wallis was used to analyze the total number of correct responses made on the first two test trials. The use of the Kruskal-Wallis was necessary because of the differ-

TABLE 2

MEDIAN AND SEMIINTERQUARTILE RANGE OF THE NUMBER OF TRIALS REQUIRED TO MAKE THE FIRST CORRECT RESPONSE TO THE THREE CRITICAL ITEMS

	Groups		
	C	E-1	E-2
Mdn.	13.00	17.00	23.00
Q	2.25	4.38	8.92

ences in the form of the three distributions and also because of heterogeneity of variance. The Kruskal-Wallis ranked analysis of variance showed that the three groups were not significantly different from each other at the .05 level ($\chi^2 = 3.55$).

The analysis of the rate of learning the three critical items was crucial to an interpretation of the data in this investigation. The response measure for each *S* consisted of the total number of trials required to make the first correct response to *all* three critical pairs. The median and semiinterquartile range of the number of trials required to make the first correct response to *all* three items are presented in Table 2.

In order to determine which of the comparisons were significantly different, the Mann-Whitney U test was used. When the C group was compared with the E-1 group, the resulting U value of 115.5 was significant at the .025 level for a one-tailed test. This comparison provides negative evidence for one-trial learning. It indicates that when the same three low-*m* stimuli and responses are retained after Test Trial 2 but are repaired, there is a significant decrement in the learning of these three S-R pairs as compared to the learning of the three pairs in the C group.

Regardless of whether one assumes the one-trial or incremental position, the E-1 and E-2 comparison provides an evaluation of the relevancy of the stimulus-differentiation variable. The resulting U value of 128.5 was significant at the .05 level and indicates a slower learning rate among the three inserted items in the E-2 group as compared to the three re-paired items in the E-1 group.

The only position in Table 1 which does not predict a difference between the C and E-2 groups is the one-trial position that considers differentiation to be irrelevant in the learning of an S-R association. This comparison yielded a U value of 76 which was significant at the .002 level and indicates a faster learning rate for the

TABLE 3

SUMMARY OF THE RELATIONSHIPS BETWEEN THE THREE GROUPS ON THE NUMBER OF TRIALS REQUIRED TO RESPOND CORRECTLY TO THE THREE CRITICAL ITEMS AND THE PREDICTED RELATIONSHIPS MADE BY THE FOUR THEORETICAL POSITIONS

Obtained Relationships	Level of Significance	Predicted Theoretical Relationships			
		Incremental		One-Trial	
		Diff. Relv.	Not Relv.	Diff. Relv.	Not Relv.
C < E-1	.025	C < E-1	C < E-1	C = E-1	C = E-1
C < E-2	.002	C < E-2	C < E-2	C < E-2	C = E-2
E-1 < E-2	.05	E-1 < E-2	E-1 = E-2	E-1 < E-2	E-1 = E-2

Note.—Diff. = Differentiation, Relv. = Relevant.

three critical items in the C group as compared to the E-2 group.

Table 3 presents a summary of the obtained relationships among the three groups on the number of trials required to respond correctly to the three critical items and the predictions made by the four theoretical positions presented in Table 1. The C and E-1 comparison provides positive evidence for an incremental development of associative strength between the three low-*m* pairs. Comparing the E-1 group with the E-2 group demonstrated the relevancy of stimulus differentiation in the formation of an S-R association. The comparison between the C group and the E-2 group demonstrated the combined effect of the subthreshold development of associative strength between the three low-*m* pairs and the lack of stimulus differentiation among the three low-*m* stimuli.

The total trials to reach criterion for the 12-item list were analyzed by the use of a 1×3 analysis of variance design. Table 4 presents the means and standard deviations of the total number of trials to reach criterion for the three groups. The $F(2, 57) = 1.66$, did not approach significance at the .05 level.

TABLE 4

MEANS AND *SD*s OF THE TRIALS TO CRITERION FOR THE C, E-1, AND E-2 GROUPS

	Groups		
	C	E-1	E-2
M	11.55	15.00	15.85
SD	7.84	8.10	7.74

Thus, in spite of the fact that the three low-*m* pairs required increasingly more trials in the C, E-1, and E-2 groups before a correct response was given, there was not a significant difference in the total number of trials to reach criterion for the three groups.

The results have been interpreted as supporting an incremental development of associative strength. They also demonstrated the importance of the stimulus-differentiation variable in learning verbal associations.

REFERENCES

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

(Received January 28, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 312–316

EVENT SALIENCE AND RESPONSE FREQUENCY IN A TEN-ALTERNATIVE PROBABILITY-LEARNING SITUATION [1]

LEE ROY BEACH

Office of Naval Research

AND RICHARD W. SHOENBERGER

Behavioral Sciences Laboratory, Wright-Patterson Air Force Base

The purpose of this investigation was to examine the influence of the salience of stimulus events on the relative frequency of response in a 10-alternative probability-learning situation. It was found that for the most salient event, differences in salience were reflected in differences in degree to which Ss' relative frequencies of response exceeded the events' relative frequency of occurrence. For the remaining less salient events this relationship did not exist.

In the two-event probability-learning situation it is commonly found that the relative frequencies of *S*s' responses closely match the events' relative frequencies of occurrence. In multiple-event situations, however, it is found that as training progresses *S*s' response frequencies "overshoot" the event frequencies for the more frequently occurring events and "undershoot" them for infrequently occurring events (Cotton & Rechtschaffen, 1958; Gardner, 1957, 1958; McCormack, 1959).

The most apparent variable which can be suggested as a possible cause of overshooting in the multiple-event situation is the number of alternative events. Indeed, in two studies Cotton and Rechtschaffen (1958) and Gardner (1957) found the degree of overshooting was greater for three alternative events than for two. Gardner (1958) found that, for from two to eight alternative events, the greater the number of alternatives the more *S*s overshot for the high-probability event. The results of these three studies are somewhat suspect in this regard, however, because of the stimulus distributions used. In all three studies one of the events has a very high relative frequency of occurrence while the remaining events were much lower and, in one experiment (Gardner, 1958), the remaining events were all of equal relative frequency. McCormack (1959) also had one high relative frequency event with the remaining alternatives equal and found overshooting of the high event. These experimental conditions raise the question of whether the number of alternative events is necessarily the important cause of overshooting. Perhaps the degree to which the relative frequency of the high-frequency event exceeds the relative frequency of the other events is of equal or greater importance. That is, perhaps the more *salient* an event is, due to its exceptionally high rate of occurrence, the more *S*s' responses overshoot its

[1] Opinions or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the views or the endorsement of either the Navy or the Air Force. The work was done while the authors were at the United States Naval School of Aviation Medicine, Pensacola, Florida. The authors would like to express their appreciation to A. C. Taylor, J. R. Berkshire, R. Balocki, and C. W. Husty for their help and advice.

objective relative frequency of occurrence. Overshooting the high-frequency events would, of course, result in undershooting the lower frequency events.

While a more complex definition may be necessary for future investigations, in the present study salience was defined as the "sharpness" of the peak in the distribution caused by an event occurring with a higher relative frequency than the events which surround it on some underlying physical or conceptual dimension. This definition permits the computation of a simple *Salience Index* value for the

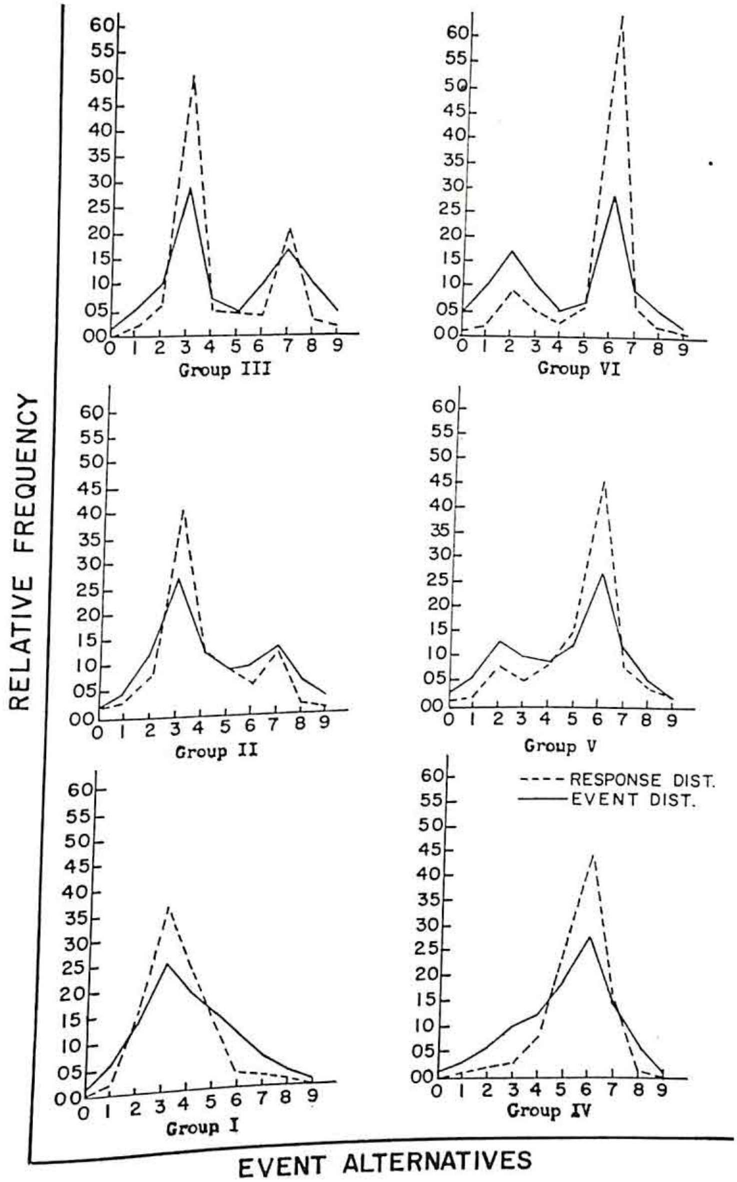


FIG. 1. Event presentation distributions and the corresponding response distributions based on Trials 1,350–1,500.

TABLE 1

Relative Frequencies of Events (E) and of Responses (R) over Last 150 Trials for Each Group of Ss

Group		Event 0	1	2	3	4	5	6	7	8	9
I	E	.01	.06	.14	.25[a]	.19	.15	.10	.06	.03	.01
	R	.00	.02	.16	.36	.24	.13	.03	.02	.01	.00
II	E	.02	.05	.12	.27[a]	.13	.09	.10	.13[a]	.06	.03
	R	.02	.03	.08	.41	.13	.09	.06	.12	.03	.01
III	E	.02	.05	.10	.29[a]	.07	.05	.10	.17[a]	.10	.05
	R	.00	.02	.06	.51	.05	.05	.04	.21	.03	.02
IV	E	.01	.03	.06	.10	.12	.19	.28[a]	.14	.06	.01
	R	.00	.01	.02	.03	.08	.25	.44	.15	.01	.00
V	E	.03	.06	.13[a]	.10	.09	.13	.27[a]	.12	.05	.02
	R	.01	.02	.08	.05	.08	.15	.46	.08	.04	.02
VI	E	.05	.10	.17[a]	.10	.05	.07	.29[a]	.10	.05	.02
	R	.01	.02	.09	.05	.03	.06	.65	.06	.02	.00

[a] High-salience events.

high-frequency events by the equation

$$SI_{E_n} = 2\Pr(E_n) - \Pr(E_{n-1}) - \Pr(E_{n+1})$$

where $\Pr(E_n)$ is the relative frequency of the high-frequency event and $\Pr(E_{n-1})$ and $\Pr(E_{n+1})$ are the relative frequencies of the events on either side of E_n on the underlying dimension. A high positive value of SI indicates a high degree of salience for Event E_n. It was predicted that *S*s' tendency to overshoot for Event E_n would be positively related to its degree of salience.

Method

The apparatus consisted of a 34 × 8 in. panel on which was mounted a row of 10 white event lights numbered from 0 to 9. Above the row of white lights was mounted an amber "ready light." The *S*s were provided with IBM Port-a-Punch boards and were instructed that when the ready light came on they were to punch into their IBM card the number of the event light they expected to come on next. (After having punched their prediction and checked its accuracy *S*s covered it with another card; this prevented them from seeing the pattern of their past responses.) The sequence was: 2 sec. ready light, 3 sec. pause for *S*s to punch their predictions, 1 sec. event light, 2 sec. intertrial pause, 2 sec. ready light again, etc. The *S*s were run for approximately 1 hr. each day until they reached a total of 1,500 trials.

Stimulus presentation orders were constructed by randomly assigning the numbers 0–9 to spaces on a list numbered from 1 to 100. Six orders were constructed for each of six different groups of *S*s. For each group of *S*s the six lists were repeated one after another. The *S*s did not know when one list ended and the next began.

The *S*s were newly commissioned Naval officers who were assigned to classes in the order of their arrival at the United States Naval Air Basic Training Command, Pensacola, Florida. While this was not a random procedure it was a close approximation. Each experimental group consisted of one class (the classes had no contact with one another) and there were six groups: 10 *S*s in Group I, 12 *S*s in Group II, 14 *S*s in Group III, 15 *S*s in Group IV, 13 *S*s in Group V, and 13 *S*s in Group VI.

Each group of *S*s was presented with a different event-frequency distribution. Three of these distributions, presented in Table 1 and graphed in Fig. 1, represent a progression in salience for Events 3 and 7, and three

represent a progression in salience for Events 2 and 6. For Group I the salience index is .17 for Event 3 and −.01 for Event 7; for Group II SI_3 = .29 and SI_7 = .10; for Group III SI_3 = .41 and SI_7 = .14; for Group IV SI_2 = −.01 and SI_6 = .23; for Group V SI_2 = .10 and SI_6 = .29; and for Group VI SI_2 = .14 and SI_6 = .41.

Results

On the graphs of the stimulus-frequency distributions in Fig. 1 are superimposed the mean proportion of *S*s giving each response averaged over the last 150 trials. These results, also given in Table 1, only partially sup-

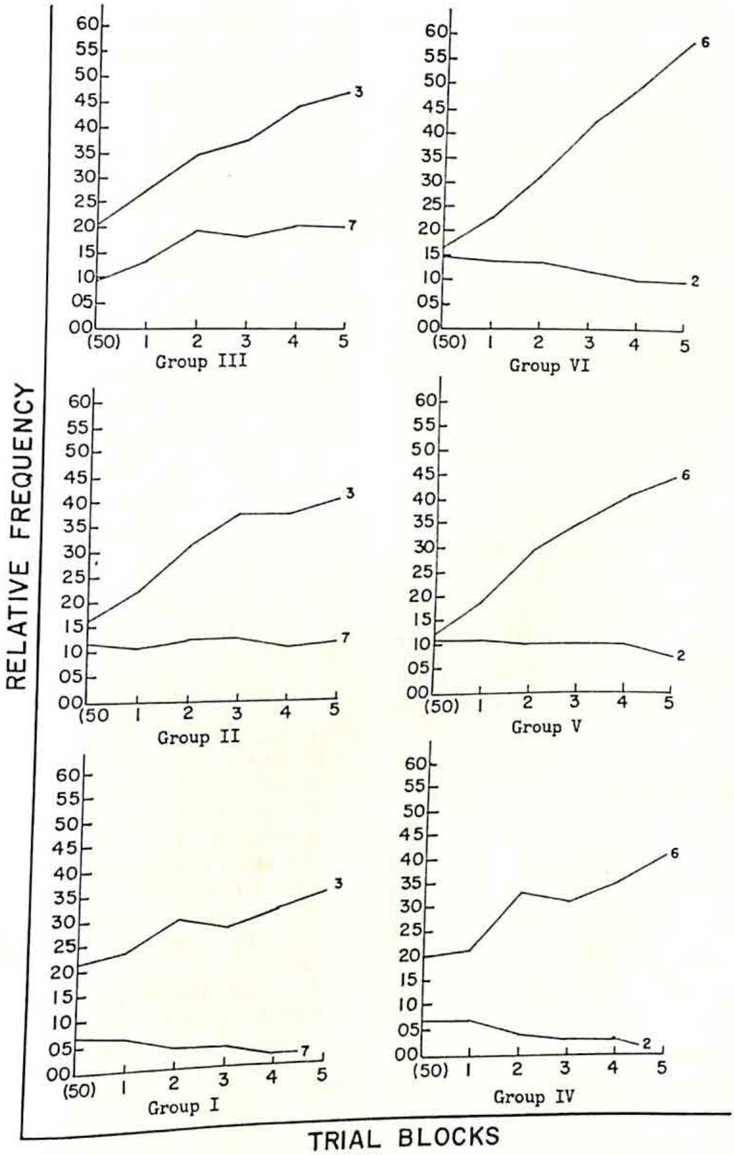


FIG. 2. Response curves for the two high-salience events. (The first point on the curve is based on 50 trials. All other points are based on blocks of 150 trials. The numbers on the right indicate the event to which the curve is related.)

port the hypothesis that overshooting is related to event salience. The increase in salience from Group I to Group III for Event 3 and from Group IV to Group VI for Event 6, is reflected in greater overshooting. (Spearman rank-order correlation between the difference between $\Pr(E_n)$ and the observed response relative frequency and SI for the high-salience event is .89, $p = .05$, $n = 6$.) However, similar increase in salience for Events 7 and 2, while reflected in high response frequencies, do not show a similar pattern of overshooting.

The graphs in Fig. 1 reveal that overshooting on the high-salience event has exercised a nonsymmetrical influence on the response frequencies for the other events. Events near the end of the 0–9 scale appear to be undershot to a greater degree than those in the middle of the scale. Moreover, events which are adjacent to the high-frequency event and which lie toward the middle of the scale, were either overshot or less undershot than the adjacent events on the other side of the high-frequency event. Clearly, these results reflect generalization effects. However the data are not sufficient to reveal the nature of these effects.

Research by Jenkins and Cunningham (1949) has shown *Ss* to have rather strong number preferences when asked to produce free-response lists of numbers from 1 through 10; the numbers 3 and 7 were the most preferred. Such preferences probably had some influence on responses in the present experiment. However, if it is assumed that all groups had roughly equivalent number preferences, the systematic differences in overshooting from group to group as a function of the salience of the high-frequency event (Fig. 1) rules out the possibility that the findings could be accounted for in terms of number preferences. In addition, the fact that salience produces overshooting for Events 2 and 6 as well as for the preferred Events 3 and 7 further reduces the possibility that number preferences cause overshooting.

The results in Fig. 1 are also reflected in the response curves in Fig. 2. These curves indicate that there was a rapid and fairly consistent increase in responses to the high-salience events throughout training, an increase which had not leveled off after 1,500 trials. The slopes of these curves generally correspond to the rank orders of the high-salience events for the various groups. On the other hand, the curves for the second most salient events leveled off early in training and subsequently even decreased for four of the six groups.

In short, the results indicate that an event's salience is not sufficient to promote overshooting. However, for the most highly salient event the degree to which overshooting occurs is positively related to the degree of salience.

REFERENCES

Cotton, J. W., & Rechtschaffen, A. Replication report: Two- and three-choice verbal conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

Gardner, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

Gardner, R. A. Multiple-choice decision behavior. *Amer. J. Psychol.*, 1958, **71**, 710–717.

Jenkins, W. O., & Cunningham, L. M. The guessing-sequence hypothesis, the "spread-of-effect," and number-guessing habits. *J. exp. Psychol.*, 1949, **32**, 158–168.

McCormack, P. D. Spatial generalization and probability learning in a five-choice situation. *Amer. J. Psychol.*, 1959, **72**, 135–138.

(Received January 31, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 317–323

DIFFERENTIAL EFFECTS OF STIMULUS AND RESPONSE ISOLATION IN PAIRED-ASSOCIATE LEARNING[1]

RAYMOND L. ERICKSON

University of New Hampshire

Either the stimulus or the response of 1 S-R pair in an 8-pair list of 2-letter consonants was isolated by printing it in red. Lists were presented either in the usual paired-associate manner or in a manner such that the stimuli and responses were never simultaneously exposed. Both stimulus isolation and response isolation produced an effect ($p < .001$), but the effect of stimulus isolation was greater ($p < .001$). Manner of presentation produced no detectable effect. Results were interpreted to be inconsistent with either (a) Underwood, Runquist, and Schulz's (1959) assumption that equal increases in stimulus and response similarity will produce equal interference effects in the associative phase of learning, or (b) an S-R intralist interference explanation for the von Restorff effect that is based solely upon reduced stimulus or response generalization.

While the more rapid learning of unique items in lists of verbal material seems well established empirically, there is no generally accepted theoretical explanation for this isolation, or von Restorff, effect (Erickson, 1963). A prerequisite to the development of such an explanation is the acquisition of additional data relevant to the properties of the isolation effect. The purpose of the present study was to gather such data by investigating more precisely the relative efficacy of isolating a stimulus as opposed to a response element in a paired-associate list. Such data are also relevant to theories of verbal learning in general, particularly those stressing the role of intralist associative interference.

Studies of differential stimulus-response isolation effects in paired-associate learning are rare, and the data available are of questionable value. Kimble and Dufort (1955), in their Exp. II, inferred that their results supported the implication that the isolation effect is a perceptual one that operates chiefly on stimuli, since they felt that they had demonstrated an isolation effect under a condition of stimulus isolation and an "inverse" isolation effect under that of response isolation. However, their design confounded the conditions of stimulus and response isolation with the kinds of material functioning as stimuli and responses. In addition, their inference was based upon a misinterpretation of the meaning of a significant F for interaction. Their significant F for interaction can only be legitimately interpreted to indicate that the relative ease of learning isolated and nonisolated pairs was *different* under the conditions of stimulus and response isolation. In order to be valid, Kimble and Dufort's interpretation that the isolated pairs were learned more easily than the nonisolated pairs when isolation was on the stimulus side would require a significant F for the simple main effect of isolation on the stimulus side. A

[1] This research was supported by Grant MH 07125-01 from the National Institutes of Health. Pilot work on the study was made possible through Grant G-12986 from the National Science Foundation.

The author wishes to thank Alan Sroufe and Leon Swartzendruber for their help in data collection.

test of this effect (Winer, 1962, p. 311), using the published data, yielded a nonsignificant result, F (1, 26) = 1.24, $p > .25$.

Nachmias, Gleitman, and McKenna (1961) attempted a more rigorous investigation of differential stimulus-response isolation effects, from which they concluded that isolation is effective on both the stimulus and the response sides of paired-associate lists, and "to about the same extent." However, the fact that they isolated either two stimulus or two response elements in each of their nine-pair lists seems unfortunate. Pillsbury and Raush (1943) have demonstrated a regular decrease in the advantage of isolated items as the ratio of disparate material to massed material is increased in serial lists. By using two isolated pairs in each list rather than one, Nachmias, Gleitman, and McKenna very likely attenuated the isolation effect, and may have thereby reduced the probability of detecting any real difference existing between the effects of stimulus and response isolation.

Erickson's (1963) subsidiary finding that isolation of stimuli and responses did not produce detectably different effects suffers from a somewhat similar limitation. In the relevant part of this study, isolation of each critical S-R pair was produced by (*a*) the unique relationship between the items constituting the critical pair, and (*b*) differential stimulus-response isolation along the dimension of color. Since any differential effects produced by stimulus vs. response isolation were in addition to the undifferentiated isolation effect produced by the unique relationship between the items of the critical pair, the fact that the data suggested a larger isolation effect under stimulus isolation than under response isolation demands further investigation.

TABLE 1

Basic Items Used in All 32 Experimental Lists

Item Position 1	Item Position 2
RV-MB[a]	MB-RV[a]
TL-CN	CN-TL
GK-SD[a]	SD-GK[a]
BJ-HG	HG-BJ
LF-VM[a]	VM-LF[a]
DH-FP	FP-DH
NR-PC[a]	PC-NR[a]
JT-KS	KS-JT

Note.—Lists were presented in eight different orders such that each pair appeared once at each of the eight positions and there was never more than one repetition of any adjacent pairs in running through eight consecutive trials of a list. The eight orders were the same for all lists.

[a] Isolated pairs. Only one pair was isolated in any given list.

The major purpose of the present study, then, was to further delineate the differential effects of stimulus and response isolation, using paired-associate lists in which the stimuli and the responses were composed of the same type of material and represented comparable levels of intralist similarity. Differential stimulus-response isolation was produced by printing either the stimulus or the response of a critical pair in red, while all other items appeared in black. The lists were otherwise identical under the two main experimental conditions.

An additional purpose was to determine the effect of altering the usual paired-associate procedure by presenting lists in such a manner that the stimuli and responses were exposed for equal amounts of time and were never simultaneously visible to *S*.

Method

Materials and apparatus.—Table 1 shows the eight pairs of basic items that were used in all lists, Item Position 2 being a mirror image of Item Position 1. Four pairs functioned as isolated pairs, only one pair being isolated in any given list. List 1 consisted of the eight pairs shown in Item Position 1,

with the stimulus item of the critical pair RV-MB printed in red. All nonisolated items appeared in black only. Lists 2–4 were identical to List 1, except that different critical stimuli (GK, LF, and NR) were isolated by being printed in red. Lists 5–8 replicated the first four, except that the *response* items were isolated along the color dimension in the above fashion.

Lists 9–16 replicated the first eight, except that the critical element of the isolated pairs was not only printed in red, but also appeared in a larger and different style type. Lists 17–32 were identical to Lists 1–16, except that they were their mirror images, consisting of the particular S-R pairs shown in Item Position 2.

The four pairs selected to function as isolated pairs were chosen because they were the pairs most comparable to each other in difficulty level in both item-position versions of the list. The basis for the decision was data from a pilot study involving 32 *S*s.

The lists were printed on a printing press on 80-lb. fotolith paper. The symbols were $\frac{11}{32}$ in. in height, printed in 36-point Clarendon Book type. The larger isolated items used on half the lists were $\frac{15}{32}$ in. in height, and appeared in 36-point Twentieth Century Extra Bold Italic. Each list appeared on an endless tape in eight different orders, with no spaces separating the orders. A Gloric memory drum with a specially constructed neon light assuring equal illumination of the stimulus and response apertures was used. The memory drum permitted two alternative exposure patterns, both based upon a 2-2-1 sec. time sequence.

Design and procedure.—In addition to the four variables specified above in the description of the 32 lists, two other independent variables were included in the design: (*a*) the particular *E* who collected the data; (*b*) stimulus exposure time. For half the *S*s, the lists were presented in the usual paired-associate manner; i.e., the stimulus was exposed for a total time of 4 sec., the last two of which were contemporaneous with the exposure of the response item. For the other half of the *S*s, the stimulus shutter closed at the end of 2 sec. just as the response shutter opened to expose the response item for 2 sec. Under the latter condition the stimulus and response items were never simultaneously visible. Before the experiment began, each of the 128 *S*s was randomly assigned to a specific condition in a completely counterbalanced factorial design involving the six variables that have been described.

Each *S* first practiced to a criterion of one perfect recitation on a list of six pairs of two-syllable nouns. He was then presented with his particular experimental list under the appropriate stimulus shutter condition. Learning continued for a total of 24 trials. The *S*s were instructed to spell out their responses and were told not to be afraid to guess and make mistakes. Responses were recorded verbatim.

A minor variable that was controlled but not counterbalanced was the particular order of the list that *S* received as his initial order of presentation. Previous work involving a similar task (Erickson, 1963) demonstrated no effect of starting order on the size of the isolation effect. In the present study, starting order was confounded with specific isolated pairs so that each *S* received a starting order that placed his particular isolated pair in Position 4 in the initial presentation of the list.

Subjects.—The *S*s were 128 volunteer undergraduate psychology students at Whittier College. There were 42 males and 86 females. One substitute was run to replace an *S* from the response-isolation condition who was rejected for failing to meet a predetermined criterion of 10 correct nonisolated responses of any kind during the 24 trials.

RESULTS

The dependent variables analyzed were: (*a*) mean number correct responses over Trials 1–24 to the three critical nonisolated pairs, each of which was used in some other list as the isolated pair; (*b*) difference between number correct responses to isolated pair over Trials 1–24 and mean number correct responses to critical nonisolated pairs over Trials 1–24; (*c*) mean number correct responses over Trials 1–24 to the four noncritical nonisolated pairs that were not used in any other list as the isolated pair. The dependent variable described under *b* was the indicator of the isolation effect. The N (nonisolated) scores, unless otherwise specified, were based only on the rate of learning the three critical nonisolated pairs described above under *a*.

Table 2 shows mean correct responses for isolated and critical non-

TABLE 2

MEAN CORRECT RESPONSES PER S PER PAIR OVER TRIALS 1–24 FOR ISOLATED AND CRITICAL NONISOLATED PAIRS UNDER CONDITIONS OF STIMULUS AND RESPONSE ISOLATION

Response Category	Isolation Cond.	
	Stimulus Isolated	Response Isolated
Isolated Pair (I)	18.58	13.52
Nonisolated Pairs (N)	9.45	8.67
Difference (I − N)	9.13	4.85

isolated pairs under the two conditions of isolation. It also shows the mean I − N score (isolated correct minus mean critical nonisolated correct), which represents the isolation effect. Figure 1 presents the learning curves for isolated and critical nonisolated pairs within each of the conditions of isolation.

As indicated by Table 2 and Fig. 1, the isolated pairs were learned more rapidly than the critical nonisolated pairs under both conditions of isolation, but the advantage of isolated over nonisolated pairs was greater under the stimulus-isolation condition. The mean difference of 4.28 in the respective I − N scores is significant, as shown by the F ratio for Stimulus vs. Response Isolation obtained in an analysis of variance of I − N scores, $F\ (1, 104) = 12.53$, $p < .001$.

An independent analysis of variance was carried out on the I − N scores obtained under the response-isolation condition only. It yielded a significant F for the mean, $F\ (1, 63) = 28.19$, $p < .001$, indicating that the mean difference between isolated correct and mean critical nonisolated correct, though significantly smaller than the corresponding difference obtained under stimulus isolation, was a significant difference.

The main analysis of variance of the I − N scores evaluated all main effects and all two-way interactions involving the Stimulus vs. Response Isolation variable. In addition, two selected two-way interactions and two three-way interactions involving the S-R Isolation variable were evaluated, the remaining interactions being pooled for error. Other than the highly significant F for the main effect of Stimulus vs. Response Isola-

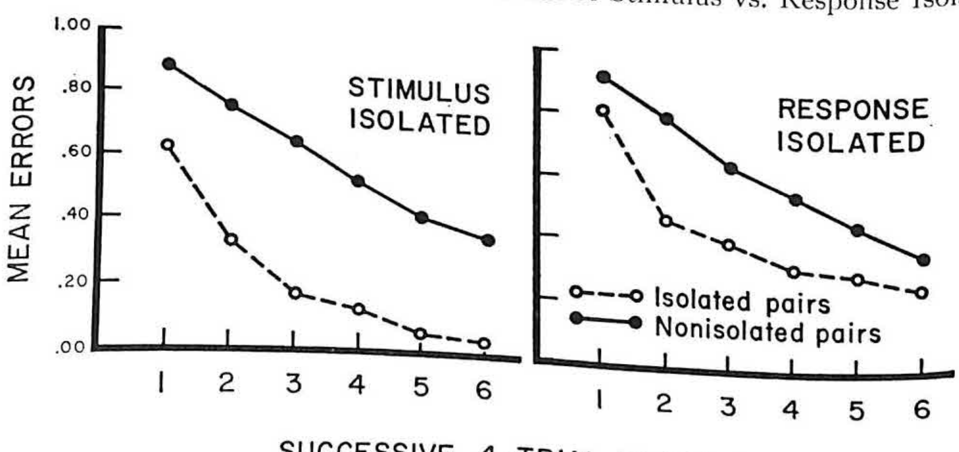


FIG. 1. Mean errors per S per pair per trial made to isolated and critical nonisolated pairs averaged over four-trial blocks for each isolation condition.

TABLE 3

Analysis of Mean Error Rate per S over Trials 1–24 for Isolated Pair under Conditions of Stimulus and Response Isolation

Isolation Cond.	Incorrect Responses to Critical Stimulus[a]						Incorrect Elicitations of Critical Response[b]
	Mean Total Errors per S	No Response	Intralist Intrusions	Partial Correct Responses[c]	Stimulus Items	Extralist Responses	
Stimulus Isolated	5.42	.60	.09	.07	.02	.22	.07
Response Isolated	10.48	.59	.18	.06	.01	.16	.14

[a] Error rate determined by dividing number of errors made in a particular category by total number of errors made by an S.

[b] Error rate determined by dividing number incorrect elicitations of critical response by a particular stimulus by total number incorrect responses made to that stimulus by an S.

[c] Partial correct responses defined as responses with one correct letter.

tion, only one significant F was obtained, that for the interaction between Stimulus vs. Response Isolation and Item Position, $F(1, 104) = 4.42$, $p < .05$. Inspection of cell means showed that: (*a*) stimulus isolation maintained an advantage over response isolation in both item positions of the list; and (*b*) in the case of every critical item, a larger I − N score was obtained when that item was isolated on the stimulus side than when it was isolated on the response side. The significant interaction reflects the fact that the advantage of stimulus over response isolation was greater when the lists were presented in Item Position 1 than when the lists were presented in Item Position 2. The cause of this difference is unknown.

An analysis of variance identical to that described above was carried out on the critical nonisolated scores. No significant Fs were obtained. The effect of Stimulus vs. Response Isolation did not approach significance, $F(1, 104) = 0.67$. One further check on possible differential rates of learning nonisolated pairs under the two main isolation conditions was made. A separate analysis of variance on the noncritical nonisolated scores (those scores based on the four nonisolated pairs in each list that did not appear in any other list as the isolated pair) was carried out. The result was nonsignificant, $F(1, 126) = 3.49, p > .05$.

Table 3 presents an analysis of the errors made to isolated pairs. Intralist intrusions (of response items from the seven nonisolated pairs) constituted a small proportion of total errors at the critical position. However, it is of interest to note that, when corrected for opportunity, the rate of intralist intrusions at the critical position and the rate of incorrect elicitations of the critical response (by the seven stimuli of nonisolated pairs) was twice as large under the response-isolation condition as under the stimulus-isolation condition. The rates of the other types of response error at the critical position were approximately the same under the two isolation conditions.

Discussion

The important findings of this investigation were: (*a*) isolation of either a stimulus item alone or a response item alone can produce the von Restorff effect in paired-associate learning; (*b*) isolation

of a stimulus item produces a greater isolation effect than isolation of a response item in paired-associate lists in which stimuli and responses represent comparable levels of intralist similarity, and; (*c*) this advantage of stimulus over response isolation is not contingent upon the longer exposure time for stimulus items characteristic of the usual paired-associate procedure.

The present results are not inconsistent with any of the major theoretical explanations for the von Restorff effect discussed by Erickson (1963). However, special attention will be given here to some of the implications of the current findings for the S-R intralist interference explanation for isolation phenomena. Such an explanation would attribute the isolation effect found here to a reduction in intralist interference, resulting from either reduced stimulus or reduced response generalization of the isolated item. If it is assumed that the interference effects resulting from stimulus generalization are greater than those resulting from response generalization, it follows that isolation of a stimulus item will produce a greater reduction in intralist interference (and hence a greater isolation effect) than a comparable degree of isolation of a response item. While the rate of intralist intrusions was small, the intrusion error data shown in Table 3 are consistent with such an explanation. The smaller isolation effect obtained under response isolation was accompanied by a rate of intralist intrusion errors that was twice as large as that obtained under stimulus isolation.

Support for the contention that the interference effects resulting from stimulus generalization are greater than those resulting from response generalization is found in the data reported by Underwood (1953a, 1953b) and Underwood, Runquist, and Schulz (1959). However, the latter investigators interpreted these same data somewhat differently. They maintained that the increased intralist interferences produced by equivalent increases in stimulus and response similarity can be assumed to have *equal* effects on the associative (as opposed to the response-learning) phase of paired-associate learning. According to their interpretation, the empirical finding that increased stimulus similarity produces greater learning difficulty than a comparable increase in response similarity would be explained by the fact that an increase in response similarity makes the response-learning phase *easier* by circumscribing what has to be reproduced, an effect not duplicated by increases in stimulus similarity.

Underwood, Runquist, and Schulz's interpretation was based upon studies of the differential effects of increased stimulus and response similarity along dimensions of either synonymity or formal letter structure. From within this context, it would be predicted that making a response item more disparate in a list would increase the difficulty of the response-learning phase for that item. This would not seem a reasonable prediction when the disparity of the response item is along the dimension of color, as it was here. There would seem to be no reason to assume that the printing of a response in red should increase the difficulty of the response-learning phase for that item; in fact, the most likely prediction would be just the opposite. If this is so, the results of the present investigation support the contention that the interference effects of a given level of stimulus similarity are greater on the associative phase of learning than are the interference effects of a comparable level of response similarity, since the same degree of differentiation produced a greater isolation effect when stimuli were involved.

The foregoing interpretation assumes an S-R intralist interference explanation based solely upon reduced stimulus or response generalization to be adequate to account for the differential isolation effect found here. Unless the effect was produced solely by a reduction in stimulus or response generalization tendencies, the present results are not necessarily relevant to Underwood, Runquist, and Schulz's assumption concerning intralist interference effects. However, if it is held that this type of S-R

explanation is adequate to account for the differential isolation effect found here, it cannot at the same time be maintained that the interferences produced by equal levels of stimulus and response similarity have equal effects on the associative phase of paired-associate learning. It is appropriate to point out that while the present data are not inconsistent with such an S-R explanation, previous studies (Erickson, 1963; Newman & Saltz, 1958; Saltz & Newman, 1959) have indicated that an S-R intralist interference explanation based solely upon reduced stimulus or response generalization is not adequate to account completely for the von Restorff effect.

A noteworthy subsidiary finding of the present study was that the alteration in stimulus exposure time used here produced no detectable effect on the ease of learning either the isolated or the nonisolated pairs. Studies comparing the relative efficacy of paired-associate and serial learning (e.g., Primoff, 1938; Young, 1959) have typically ignored a substantial procedural difference between the two tasks. In the usual paired-associate task, stimuli are exposed twice as long as responses, which are never exposed without their appropriate stimuli also being visible. Neither of these conditions obtain in the typical serial-learning task. The variation in paired-associate procedure used here, in which the stimulus shutter closed just as the response shutter opened, more closely approximates the typical serial-learning task. The finding that this alteration in procedure produced no detectable effect in paired-associate learning is consistent with the contention that the described difference between serial and paired-associate learning tasks is not an important one. However, a more thorough study of the parameters of this procedural difference, using both serial and paired-associate tasks, is called for.

REFERENCES

Erickson, R. L. Relational isolation as a means of producing the von Restorff effect in paired-associate learning. *J. exp. Psychol.*, 1963, **66**, 111–119.

Kimble, G. A., & Dufort, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50**, 361–368.

Nachmias, J., Gleitman, H., & McKenna, V. The effect of isolation of stimuli and responses in paired associates. *Amer. J. Psychol.*, 1961, **74**, 452–456.

Newman, S. E., & Saltz, E. Isolation effects: Stimulus and response generalization as explanatory concepts. *J. exp. Psychol.*, 1958, **55**, 467–472.

Pillsbury, W. B., & Raush, H. L. An extension of the Koehler-Restorff inhibition phenomenon. *Amer. J. Psychol.*, 1943, **56**, 293–298.

Primoff, E. Backward and forward association as an organizing act in serial and in paired associate learning. *J. Psychol.*, 1938, **5**, 375–395.

Saltz, E., & Newman, S. E. The von Restorff isolation effect: Test of the intralist association assumption. *J. exp. Psychol.*, 1959, **58**, 445–451.

Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142. (a)

Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149. (b)

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Young, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

(Received February 7, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 324–325

SUPPLEMENTARY REPORTS

FURTHER EFFECTS OF SUBJECT-GENERATED RECODING CUES ON SHORT-TERM MEMORY[1]

RICHARD H. LINDLEY
Eastern Michigan University

AND

SHARI E. NEDLER
Trinity University

48 men and women college students were tested in a short-term memory (STM) study using the Peterson and Peterson (1959) technique. All Ss were tested with 4 retention intervals, high- or low-meaningfulness (HM or LM) items, and 2 presentations of the trigram or recoding cue preceding or following the trigram. 24 Ss recalled both the trigram and the recoding cue; 24 recalled only the trigram. Major results were (a) recoding cues in either order facilitated trigram recall; (b) recalling both the trigram and the recoding cue had no effect on trigram recall as compared to recalling the trigram alone; (c) correct recall for LM recoding cues was 93%, for HM cues 92%.

This study follows up a study by Schaub and Lindley (1964). Short-term memory (STM) for trigram-recoding cue pairs was explored as a function of the two possible orders of item presentation (e.g., HONEY-HON and HON-HONEY) and of two different recall tasks (recall trigram or recall trigram and recoding cue).

Method.—The $4 \times 2 \times 3 \times 2$ factorial design had four retention intervals (0, 8, 14, 20 sec.), either high-meaningfulness (HM) or low-meaningfulness (LM) trigrams, three types of item presentations (two presentations of trigram, recoding cue following trigram, recoding cue preceding trigram), and two types of retention task (recall trigram and recall trigram plus recoding cue). All *S*s were tested with the same retention intervals, levels of meaningfulness, and types of item presentation; half the *S*s were instructed to recall the trigram alone and the other half were instructed to recall both the trigram and the recoding cue in the order that they thought of them.

One random order of the 48 items was used. The *S*s had 12 sec. to recall the items and a 6-sec. rest was given after the recall period. The items were the 24 LM and 24 HM trigrams of Schaub and Lindley (1964) with the corresponding common recoding cues. Forty-eight men and women paid volunteers from Trinity University served as *S*s; they were assigned to the two conditions of recall task alternately. In all other respects the procedure used by Schaub and Lindley (1964) was followed.

Results.—Table 1 presents the proportions of trigrams recalled correctly. The analysis of variance of these data (excluding the 0-sec. retention interval) confirmed the Schaub and Lindley (1964) study in essential details; meaningfulness, retention intervals, and the presence of recoding cues were all significant sources of variance, F (1, 782) (2, 782) (1, 782) = 143.82, 6.63, 39.55, $p < .01$, respectively. The order of trigram-recoding cue presenta-

[1] This research was supported by a grant to the first author from the National Science Foundation.

TABLE 1
PROPORTIONS OF TRIGRAMS RECALLED CORRECTLY

	Recall Trigrams Only								Recall Trigrams and Recoding Cues							
	Retention Intervals in Sec.								Retention Intervals in Sec.							
	LM				HM				LM				HM			
Cond.	0	8	14	20	0	8	14	20	0	8	14	20	0	8	14	20
Item Presentation																
Trigrams twice	.98	.69	.58	.54	1.00	.81	.71	.71	1.00	.67	.48	.50	1.00	.83	.81	.67
Trigram-Cue	.96	.65	.62	.65	.98	.96	.94	.83	.98	.65	.65	.67	1.00	.96	.87	.87
Cue-Trigram	.96	.69	.79	.65	1.00	.83	.87	.83	.96	.67	.71	.69	1.00	.96	.92	.83

TABLE 2

TRIGRAM AND RECODING CUE RECALL: PROPORTION OF TOTAL RESPONSES IN EACH CATEGORY

	LM			HM		
	Trigram Correct	Trigram Incorrect	Totals	Trigram Correct	Trigram Incorrect	Totals
Recoding cue correct	.65	.28	.93	.90	.02	.92
Recoding cue incorrect	.02	.05	.07	.00	.08	.08
Totals	.67	.33	1.00	.90	.10	1.00

tion was not significant, $F\ (1, 782) < 1$; thus both of the trigram-recoding cue sequences facilitated recall. The sequence: trigram followed by recoding cue presumably mimics the normal sequence in which *S* searches for a recoding cue after the trigram has been presented. Inspection of Table 1 also shows that trigram recall is not affected by requiring the recall of the recoding cue along with the trigram, $F\ (1, 24) < 1$; if *S*s use the recoding cue to generate the trigram, then it should be immaterial whether or not the overt recall of the recoding cue is required.

The only other significant differences were Retention Intervals × Recoding Cues, $F\ (1, 782) = 3.92$, $p < .05$, and Meaningfulness × Order of Item Presentation, $F\ (2, 782) = 3.73$, $p < .05$. The former interaction is significant because the rate of forgetting of trigrams with recoding cues present is slower than the rate when the trigram is presented twice; the latter interaction is significant because the trigram-recoding cue sequence leads to superior recall with HM items and the recoding cue-trigram sequence is superior with LM items.

The 24 *S*s who recalled both the trigram and the recoding cue aloud were instructed to recall them in the order that they thought of them; there were 118 out of a possible 576 responses (excluding the 0-sec. retention interval) in which the order of recall was the reverse of the order of presentation. If *S* uses the recoding cue to generate the trigram, then the order of recall: *recoding cue preceding trigram* should be a more frequent type of reversal than the order of recall: *recoding cue following trigram.* This was the case; the mean number of the former type of reversals per *S* was 3.46, of the latter, 1.46, $t\ (22) = 4.00$, $p < .01$.

Table 2 presents the data for the four possible combinations of correct or incorrect trigram recall and correct or incorrect recoding cue recall. The level of recall of the LM and HM recoding cues is almost identical and most of the errors for the LM trigrams occur although the recoding cues have been correctly recalled. Presumably these two findings are due to the degree to which the first three letters of the recoding cue are in the same sequence as the letters of the trigrams. That is, for the typical LM combination (e.g., ZUJ-ZULU), if *S* recalls the recoding cue and spells the first three letters of it, he will make a mistake. On the other hand, for the typical HM combination (e.g., HON-HONEY), a similar strategy will lead to a correct recall. Thus the relatively poor recall of the LM trigrams appears to be a function of the ease with which the trigram can be generated from the recoding cue. The error data provide a further test of this hypothesis.

"Decoding errors" were defined as three-letter errors in which the incorrect letter or letters of the error were present in the recoding cues; e.g., ZUL for ZUJ-ZULU. The three-letter errors from all 48 *S*s were classified as either decoding errors or "other" errors; for the LM pairs there were 91 decoding errors and 79 other errors; for the HM pairs the corresponding figures were 2 and 21, respectively. (For comparison, the number of errors classified as decoding errors from the conditions where the trigram was presented twice were 5 for LM trigrams and 3 for HM trigrams; the corresponding other errors in these conditions were 69 and 31, respectively.) Thus the largest number of decoding errors are found in the condition where decoding the trigram from the recoding cue should be most difficult; i.e., where the same order of letters is not present in both the trigram and the recoding cue. Finally, since the level of recall for the LM recoding cues is as high as for the HM cues, the difference in level of LM and HM trigram recall *under the present conditions* may be solely a function of the degree to which the rules for generating the trigram from the recoding cue are of an equal degree of complexity.

REFERENCES

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, 58, 193–198.

SCHAUB, G. R., & LINDLEY, R. H. Effects of subject-generated recoding cues on short-term memory. *J. exp. Psychol.*, 1964, 68, 171–175.

(Received January 21, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 326–327

FIRST-LIST RETENTION AS A FUNCTION OF THE METHOD OF RECALL [1]

JOHN P. HOUSTON,[2] BERTRAM E. GARSKOF,[3] DALE E. NOYD,[4] AND JANICE M. ERSKINE [5]
University of Michigan

The present study compared amounts of retroactive inhibition (RI) obtained with the MMFR technique (Barnes & Underwood, 1959) and a single-list recall method. Following A-B, A-C learning Ss recalled either the responses from both lists or the responses from the 1st list. The results indicated that both recall methods resulted in significant RI but that the amounts of RI did not differ.

In a recent study of first-list retention in the A-B, A-C and A-B, A-B′ paradigms, Postman (1962) found that the conventional anticipation method of measuring List 1 recall resulted in measures of retroactive inhibition (RI) which did not differ from those obtained with the MMFR technique developed by Barnes and Underwood (1959). Since conventional anticipation of List 1 responses involves competition from List 2 responses, while the MMFR technique presumably eliminates this factor, Postman interpreted the similarity between the amounts of RI as measured by the two methods to mean that RI immediately following interpolated learning must be determined principally by List 1 associative strength and that response competition has little effect upon the magnitude of RI. The present experiment was designed to supplement this finding by comparing amounts of RI obtained with the MMFR technique and a single-list recall method. Following A-B, A-C learning, *S*s recalled either the responses from both lists or the response from the first list.

Method.—Three groups of 27 *S*s learned two successive lists conforming to the A-B, A-C paradigm. A fourth group of 27 *S*s learned only an A-B list. The A-B and A-C lists employed in the study were those used by Barnes and Underwood (1959). They were composed of eight nonsense syllables as stimuli and two sets of eight two-syllable adjectives as responses. The lists possessed no apparent interlist or intralist similarity. The three pairings of the stimuli and responses used by Barnes and Underwood were also employed in the present study. Nine *S*s in each condition were given each of the three pairings. An equal number of *S*s in each group received each of the two lists as first and second lists. Three random orders of the pairs in each list were used to minimize serial learning. The four conditions were randomized 27 times so that in 27 blocks of four *S*s each condition occurred only once. The *S*s were assigned to these blocks in the order of their appearance in the laboratory. All lists were presented on a Stowe memory drum at a 2:2 sec. rate with a 4-sec. intertrial interval. The *S*s in three of the four conditions learned the A-B and A-C lists to a criterion of one perfect trial. A 1-min. interval separated the two lists. A Control group learned the A-B list to a criterion of one perfect trial and then engaged in a maze-tracing task for 10 min., the approximate time taken by the remaining groups to learn the A-C list. Following A-C learning, the MMFR Group *S*s were presented the stimuli on the drum and asked to recall both first- and second-list responses in the order they came to mind. The First-List Recall group was instructed to recall and say aloud only the List 1 responses. Second-List Recall group recalled only List 2 responses. Following the maze task the Control group *S*s were asked to recall the A-B responses when presented with the stimuli. The three stimulus orders used during learning were presented equally frequently in all groups during the recall trial. Each stimulus was presented for 12 sec. during the test trial.

Results.—The mean numbers of trials to a criterion of one perfect trial on the A-B list were 10.37, 10.78, 10.96, and 11.11 for the MMFR, First-List Recall, Second-List Recall, and Control groups, respectively. The F for these values was not significant, $F\ (3, 104) = .05$, indicating that the four groups were comparable in terms of learning ability.

The mean number of correct A-B responses recalled by the MMFR group was 4.44. The mean for the First-List Recall group was 4.48

[1] This work was supported in part by Research Grant No. MH-04239 from the National Institutes of Health, United States Public Health Service, to Edward L. Walker.

[2] Now at University of California, Los Angeles.

[3] Now at Rutgers University.

[4] Now at United States Air Force Academy.

[5] Now at American Institutes of Research.

and for the Control group 7.88. An analysis of variance applied to these means indicated a significant difference between them, $F(2, 78) = 49.54$, MS error $= 2.13$, $p < .001$. The difference between the MMFR and the First-List Recall means was not significant, $t(52) = .08$. These results indicate that significant RI was demonstrated in both the MMFR and the First-List Recall conditions but that the amount of RI as measured by these two techniques did not differ.

Mean A-C recall for the MMFR group was 7.74 while mean A-C recall for the Second-List Recall group was 7.70. The difference between these means was not significant, $t(52) = .21$.

It is worth noting that the data of the present study indicate the reliability of the Barnes and Underwood (1959) findings. The A-B and A-C recall means for both the MMFR and the First- and Second-List Recall groups closely parallel the means obtained by Barnes and Underwood at high levels of second-list learning. This agreement is particularly interesting in the light of the fact that the present study involved fixed recall intervals rather than *S*-paced recall. The implication is that *S*-pacing procedures may be of little importance in relation to the unlearning effect.

REFERENCES

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Postman, L. Retention of first-list associations as a function of the conditions of transfer. *J. exp. Psychol.*, 1962, **64**, 380–387.

(Received January 27, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 327–328

EFFECT OF EXTRANEOUS STIMULATION ON THE VISUAL PERCEPTION OF VERTICALITY: A FAILURE TO REPLICATE [1]

ROBERT FRIED
Educational Testing Service

AND

RICHARD G. LATHROP
Douglass College

This study was a partial replication of a previous study of the visual perception of the vertical. Under conditions of unilateral auditory stimulation, and with the head held firmly in place, 20 Ss adjusted a line segment until it appeared vertical. A previously reported shift in apparent vertical was not obtained. It is suggested that change in body attitude may be responsible for changes in apparent vertical under conditions of unilateral stimulation.

The current study is a partial replication of a study by Werner, Wapner, and Chandler (1951) which purported to show that direct auditory stimulation influences visual perception of verticality. Werner et al. report that female *S*s demonstrate considerably less distortion in their perception of the vertical than do male *S*s. It seemed possible that female *S*s might exhibit a reliable distortion of apparent vertical if the experimental procedure were more carefully controlled. Therefore, instead of requiring *S*s to relay their judgments of the vertical to *E*, the method of adjustment was used in the current study. Also, Werner et al. required *S*s to wear goggles between trials, to remove them to make judgments, and then to replace them between trials. To eliminate this unstable condition of adaptation, all *S*s in the current study performed without goggles in a darkened room to which they had become adapted. Finally, to eliminate all possibility of head movement, a metal headrest extending 2½ in. beyond the forehead was utilized.

Method.—Twenty female undergraduate students were assigned at random to one of the four factorial conditions of right or left auditory stimulation and control-experimental order. Ten trials were administered in both experimental and control conditions. The control condition consisted of trials during which no auditory stimulation was presented.

The *S*s were required to adjust a luminous line segment by means of a knob until it ap-

[1] The present study was supported by a grant from the Personality Research Group, Educational Testing Service, Princeton, New Jersey.

peared to be vertical. The line segment, 1 mm. wide and 6 cm. long, was 48 in. from *S*. Headrest apertures permitted *S*s to look into the apparatus and perceive the line segment. The only source of illumination in the experimental room was a dim light in the back of the apparatus. This light illuminated a rod (connected to the line segment disk) which served to indicate angular displacement. This displacement was measured to the nearest 0.25°. Auditory stimulation of .15 v. at 1,000 cps was presented through two matched stereo headphones (8 ohms impedance).

The method of adjustment was employed. Alternating ascending and descending series were utilized. The *S*s were permitted to overshoot and retrace until satisfied with their judgment of apparent vertical. Each *S* was given five practice trials to insure familiarity with the apparatus and to permit adaptation to conditions of ambient light (15-min. adaptation time).

Results.—A control point of subjective verticality (PSV_c) was computed for each *S*. For the experimental condition, an auditory stimulus point of subjective verticality (PSV_a) was computed. A difference score, *d*, was computed for each *S* ($d = PSV_c - PSV_a$). On the basis of the Werner et al. study, it was expected that this score would differ in sign for those *S*s receiving left-ear stimulation as contrasted to those *S*s receiving right-ear stimulation. Consequently, all *d* values for the left-ear stimulation condition were multiplied by -1.0. Twenty *d* scores, representing mean differences between control and auditory stimulation, corrected for *vector sign*, were obtained. A simple *t* test was computed for these data. The mean of the difference scores ($\bar{d} = .11°$) was not significantly different from zero, $t\ (19) = .22$.

The possibility remained that procedural changes were responsible for increases in *within-S*s variability. However, the control vs. stimulation correlation of .68 for the present study reflects a decrease, rather than an increase in *within-S*s variability. Thus the sensitivity of the present study to small systematic differences is greater than for the Werner et al. study.

Discussion.—The inability to obtain consistent shifts in visual perception of the vertical under conditions of unilateral auditory stimulation requires explanation. It is possible that the unifying factor underlying the Werner et al. studies is related to a shift in the physical attitude of the body. In the present study, the phenomenon is no longer observed when the head is fixed during judgments of verticality. Other procedural changes of the current study may also have contributed to the lack of a systematic shift.

REFERENCE

WERNER, H., WAPNER, S., & CHANDLER, K. A. Experiments on sensory-tonic field theory of perception: I. Effect of extraneous stimulation on the visual perception of verticality. *J. exp. Psychol.*, 1951, 42, 341–345.

(Received February 4, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 3, 329–330

Beginning with the first issue of Volume 69 and continuing through the remainder of that Volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are the entire backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication. The articles listed below are scheduled to appear approximately 11 months hence.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Effects of Intertrial Reinforcement on Resistance to Extinction Following Extended Training: Roger W. Black and Kenneth W. Spence*: Department of Psychology, University of Texas, Austin 12, Texas.

Information and Incentive Value of the Reinforcing Stimulus in Verbal Conditioning: Charles D. Spielberger,* Ira H. Bernstein, and Richard G. Ratliff: Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

Detection in Metacontrast: Peter H. Schiller* and Marilyn C. Smith: Department of Psychology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Aftereffects and Delay of Reward: E. J. Capaldi* and Hugh Poynor: Department of Psychology, Mezes Hall 211, University of Texas, Austin 12, Texas.

Short-Term Memory With a Guessing Technique: James V. Hinrichs*: Department of Psychology, Stanford University, Stanford, California.

A New Illusory Effect of the Müller-Lyer Figure: Paul T. Mountjoy*: 2422 Kensington Drive, Kalamazoo, Michigan.

The Serial-Position Effect of Ordered Stimulus Dimensions in Paired-Associate Learning: Sheldon M. Ebenholtz*: Department of Psychology, Connecticut College, New London, Connecticut.

Some Effects of Context on the Slope in Magnitude Estimation: F. Nowell Jones* and Morris J. Woskow: Department of Psychology, University of California, Los Angeles, California 90024.

Unlearning in Serial Learning: Geoffrey Keppel*: Department of Psychology, University of California, Berkeley, California 94720.

Effects of Reinforcement Delay during Learning on the Retention of Verbal Material in Adults: Larry M. Lintz* and Yvonne Brackbill: Department of Psychology, University of Denver, University Park, Denver, Colorado 80210.

Effect of a Composite Instructional Set on Responses to Complex Sounds: Stanley J. Rule* and John W. Little: Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

Serial Position as a Cue in Learning: The Effect of Test Rate: Slater E. Newman*: Department of Psychology, North Carolina State College, Raleigh, North Carolina.

Intermediate Size Discrimination in Seven- and Eight-Year-Old Children: Michael D. Zeiler* and Ann M. Gardner: Psychological Laboratory, Pendleton Hall, Wellesley College, Wellesley, Massachusetts 02181.

Verbal Transfer as a Function of S_1-R_2 and S_2-R_1 Interlist Similarity: John P. Houston*: Department of Psychology, University of California, Los Angeles, California 90024.

A Comparison of Stimulus Generalization Following Variable Ratio and Variable Interval Training: David R. Thomas* and Richard W. Switalski: Department of Psychology, Kent State University, Kent, Ohio.

Visual Field Position and Word-Recognition Threshold: Willis Overton and Morton Wiener*: Department of Psychology, Clark University, Worcester 10, Massachusetts.

* Asterisk indicates author for whom address is supplied.

Muscular Effort and Electrodermal Responses: Lawrence A. Pugh, Carl R. Oldroyd, Thomas S. Ray,* and Mervin L. Clark: University of Oklahoma Medical Center, P. O. Box 151, Norman, Oklahoma.

Verbal Repetition and Connotative Change: Harriett P. Amster* and Lynette Diamond Glasman: Institute of Human Learning, University of California, 2241 College Avenue, Berkeley, California.

Relation between Response Amplitude and Reinforcement: Dale R. Williams*: Department of Psychology, University of Pennsylvania, 106 College Hall, Philadelphia, Pennsylvania 19104.

Keeping Track of Sequential Events: Effects of Rate, Categories, and Trial Length: Richard A. Monty,* Harvey A. Taub, and Kenneth R. Laughery: Cornell Aeronautical Laboratory, Inc., P. O. Box 235, Buffalo 21, New York.

Journal of Experimental Psychology

Vol. 69, No. 4 April 1965

IMPROVEMENT OF VISUAL AND TACTUAL FORM DISCRIMINATION [1]

ANNE D. PICK

Macalester College

A discrimination-learning situation and subsequent transfer tests were used to investigate 2 hypotheses about improvement in discrimination: a "schema" hypothesis and a "distinctive feature" hypothesis. 1 visual and 2 tactual discrimination experiments were conducted. Results suggested the superiority of the distinctive feature hypothesis, at least under conditions of a simultaneous comparison, for accounting for children's improvement of discrimination of the letter-like forms used as material.

Gibson, Gibson, Pick, and Osser (1962) demonstrated that children between the ages of 4 yr. and 8 yr. improve in their ability to make visual discriminations among letter-like forms. The present study sought to determine if some kind of learning can produce such improvement in discrimination. The first experiment reported here explored this question with respect to visual discrimination, and the second and third experiments extended the investigation to tactual discrimination.

Two general hypotheses about the nature of learning during improvement of discrimination can be identified. One can be loosely termed a "schema" hypothesis, and is suggested in discussions of Bruner (1957a, 1957b), Vernon (1952, 1955), and in a recent book on perceptual development by Solley and Murphy (1960). Although these investigators deal primarily with identification behavior, i.e., recognition and categorizing behavior, their discussions implicate discrimination behavior as well. According to this point of view, discrimination and identification involve matching sensory data or "cues" about objects to prototypes or models of the objects which have been built

[1] This research was part of a doctoral dissertation at Cornell University and was supported in part by an interdepartmental grant to Cornell University from the United States Office of Education, Department of Health, Education, and Welfare. The author is grateful to Eleanor J. Gibson under whose direction the research was conducted and to Paul Jenson and Herbert L. Pick, Jr. for critically reading the manuscript. The principals and teachers of South Hill School in Ithaca, New York, and University of Minnesota Elementary School, Tuttle School, Marcy School, and Pratt School in Minneapolis, Minnesota were generous in their cooperation.

up through repeated experience with the objects and "stored" in memory. Improvement in discrimination would involve first constructing schemata or models of the objects to be discriminated, and then matching the sensory data to the models so as to identify them as "same" or "different." Practice with objects to be discriminated, then, would enable S to build up and refine the appropriate schemata.

A second general hypothesis about improvement in discrimination, a "distinctive feature" hypothesis, is suggested by the Gibsons and their colleagues (Gibson & Gibson, 1955; Gibson et al., 1962). It utilizes the concept of distinctive features introduced by Jacobson and Halle (1956) in a discussion of phoneme characteristics. Distinctive features can be thought of as dimensions of difference which distinguish and provide contrasts among objects. The hypothesis developed from the Gibsons' work is that improvement of discrimination consists of learning the distinctive features of the objects to be discriminated. The function of practice, according to this point of view, is to enable S to respond to an increasing number of stimulus variables and to discover which of these variables are "critical" in the sense that they serve to distinguish between one object and another.

The physical conditions resulting in improvement of discrimination do not *necessarily* differ for these two hypotheses. Both would predict that such improvement will occur as a function of practice with the objects to be discriminated. Hence a transfer design can be employed to determine the extent to which prototype learning, distinctive feature learning, or both will occur during training. Specifically, Ss could be presented with initially undifferentiated stimulus forms and trained to discriminate among them. In a transfer task they could then try to discriminate among stimulus forms which *either* differ from each other in the same dimensions as those which they learned to discriminate *or* which have the same prototypes as those which they learned to discriminate. Differential performance in these two transfer conditions should shed light on the function of prototype learning and distinctive feature learning in improvement of discrimination. Since the two processes are not mutually exclusive, relative transfer in the two conditions described can be determined by comparing them to a control condition in which both the prototypes and dimensions of difference of the forms differ from those used in training.

Experiment I

Method

Subjects

The Ss were 60 kindergarteners. Each S was randomly assigned to one of three equal transfer groups. There were an approximately equal number of boys and girls among the 20 Ss in each group.

Materials

The stimulus forms were letter-like forms of the kind used in the developmental study of Gibson et al. (1962). The forms, approximately 1×1 in., were black and were mounted on rectangular white cardboard cards 3×4 in.

There were six standard forms and six different transformations of each of the standard forms. These transformations were one change of a straight line to a curve, two changes of straight lines to curves, a right-left reversal, a 45° rotation, a perspective transformation equivalent to a 45° backward tilt, and a 25% increase in size. The standard forms and six transformations for each are shown in Fig. 1.

Procedure

Training.—The training procedure was the same for all Ss. The S was seated in front of

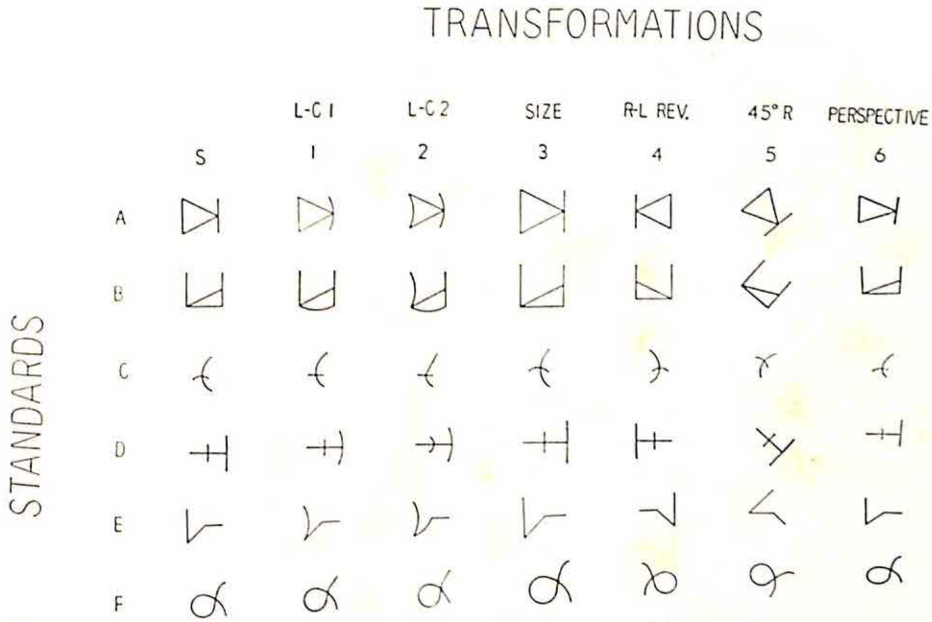


FIG. 1. Standards and transformations used in Exp. I.

a small table on which a stand similar to a lectern was placed. Three standard forms were placed on this stand and a pack of cards containing two copies of each standard and three transformations of each standard was spread out in front of *S* on the table. The *S* was instructed to look carefully at each of the 15 cards and decide whether it was exactly the same as one of the standards or if it was different. When *S* found one which was exactly the same as one of the standards, he was instructed to give it to *E*. When *S* had made a judgment about every card in the pack (i.e., finished one trial), *E* shuffled the cards and the procedure was repeated until *S* reached a criterion of one perfect trial, i.e., gave *E* only the two copies of each standard. Confusion errors (transformations which *S* indicated were exact copies of a standard) were recorded. A correction procedure, in which *E* told *S* whether each judgment was right or wrong, was used on every trial except the first.

Before the first training trial, a pretraining practice trial with real letters and correction procedure was carried out in order to acquaint *Ss* with the task and to ensure that they understood that only forms which were "exactly the same" as the standard should be given to *E*.

Transfer.—Following the criterion training trial, the transfer procedure was carried out. The task was the same as in training but only one trial was given and there was no correction given. Confusion errors were recorded as before.

The *Ss* were divided into three transfer groups in a predetermined arbitrary order. These groups differed in terms of the particular forms used in the transfer trial. Group C provided a base line with which to compare the transfer performance of the other two groups. This group received three standards and three transformations all of which were different from the ones used in training. For example, if in training these *Ss* had learned to distinguish copies of Standards A, B, and C from line to curve and size transformations of these standards, then their transfer task was to distinguish copies of Standards D, E, and F from rotation, reversal, and perspective transformations.

Group EI reflected the extent to which standard or prototype learning had occurred during training. This group received the same standards as in training, but three new transformations of these standards. For example, if in training these *Ss* learned to distinguish Standards A, B, and C from line to curve and size transformations of these standards, then their transfer task was to distinguish these *same* standards from reversal, rotation, and perspective transformations.

Group EII reflected the extent to which distinctive feature learning had occurred during training. This group received three new standards, but the same three types of trans-

formations of these standards as those with which they dealt during training. For example, if in training these *S*s learned to distinguish Standards A, B, and C from among reversal, rotation, and perspective transformations of these standards, then, in the transfer trial, their task was to distinguish Standards D, E, and F from reversal, rotation, and perspective transformations of these standards.

In order to balance the design for possible differences in difficulty of discriminating specific combinations of standards and transformations, four subgroups of *S*s were used in the training condition. Each was trained with a different combination of standards and transformations. One had Standards A, B, and C with line to curve and size transformations. Another had Standards D, E, and F and these same transformations. A third subgroup was trained with Standards A, B, and C and reversal, rotation, and perspective transformations, and a fourth had Standards D, E, and F and these transformations.

There were also, of course, 4 subgroups within each of the 3 transfer groups since the combination of forms used in transfer for a given *S* depended on the combination of forms used in training. Thus there were 12 transfer subgroups with five *S*s in each.

Results

Training

Differences between the means of the groups in number of confusion errors made on the first trial and in number of trials to criterion were analyzed using *t* tests. No differences approached the .05 level of significance and hence the null hypothesis that these groups were from the same population could be accepted.

Transfer

Confusion errors in the transfer trial constituted the main data of the experiment. Table 1 shows the errors made in the transfer trial by each of the 12 subgroups and for the three transfer groups with subgroups combined.

An analysis of variance was performed on the data for the 12 subgroups. Only the effect of transfer groups was significant, $F\ (2, 48) = 12.69$, $p < .001$. Thus the differences among the three experimental groups obtain regardless of subgroups, i.e., regardless of the particular combination of standards and transformations used.

The differences between the three groups in transfer errors were analyzed with *t* tests. The three groups were significantly different from each other with a probability level of less than .01 with a two-tailed test.

TABLE 1

CONFUSION ERRORS IN TRANSFER TRIAL FOR THE 12 SUBGROUPS

Group	Transformations				Total
	T 123		T 456		
	Standards				
	SABC	SDEF	SABC	SDEF	
EI	12	21	17	19	69
EII	11	6	10	12	39
C	23	21	32	25	101

Note.—Column headings indicate the particular combination of standards and transformations used in the transfer trial. $n = 5$ in each subgroup.

Discussion

The results of this experiment suggest that learning distinctive features may be a significant component of improvement in visual discrimination of letter-like forms. The *S*s who in the transfer trial dealt with forms which they had never seen before but which varied from each other in familiar ways (the EII group) made the fewest confusion errors. This suggests that during the training trials *S*s were learning how the forms varied from each other as they improved in their ability to discriminate among them.

The fact that Group EI, the group having familiar standards and new transformations, was superior to the control group suggests that prototype learning

also occurred during training. However, the clear superiority of Group EII implies that such learning may not be essential to improvement in discriminations of this sort.

Experiment II

The purpose of this second experiment was to investigate the generality of the results of the previous experiment for improvement in tactual discrimination. Adaptations of the procedure and materials of the previous experiment were made in order to provide appropriate conditions for studying improvement in tactual discrimination. These adaptations are noted below. The basic method and design of this experiment were the same as in the previous one.

Method

Subjects

The *Ss* were 72 first graders. First-grade children were used as *Ss* because pilot work indicated the task was better suited to this age group than to the kindergarteners used previously.

Materials

The stimulus forms were metal reproductions of some of the letter-like forms used in the previous experiment. The forms (1 × 1 in.) were made by an engraving process and were raised lines on a smooth square metal background about $1\frac{3}{4} \times 1\frac{3}{4}$ in.

There were four standard forms and 10 transformations of each: one, two, and three changes of lines to curves, a 25% increase in size, two topological transformations: break and close, a right-left reversal, 45° and 90° rotations, and a perspective transformation equivalent to a 45° backward tilt. The standard forms and their transformations are shown in Fig. 2.

Procedure

Training.—The *S* was seated at a small table on which a form board was placed. This plywood board 12 × 15 in. with a raised block in the middle, was used to display the forms in front of *S*. The raised block was covered with "velcro" as were the backs of the metal forms and the forms could thus be made to adhere to the block for presentation to *S*. The *S* was blindfolded and a standard form and one of its transformations were placed on the board. The standard was always presented on the left and the transformation on the right.

The *S*'s dominant hand was placed on the board and he was instructed to feel first one form and then the other with that same hand and decide whether the two forms were the same or different. When *S* had made a judgment about the pair of forms, the transformation was removed and replaced by another. After a few presentations, *E* no longer had to guide *S*'s hand to the board. A trial consisted of seven presentations for comparison: one standard form to be compared with each of five transformations and

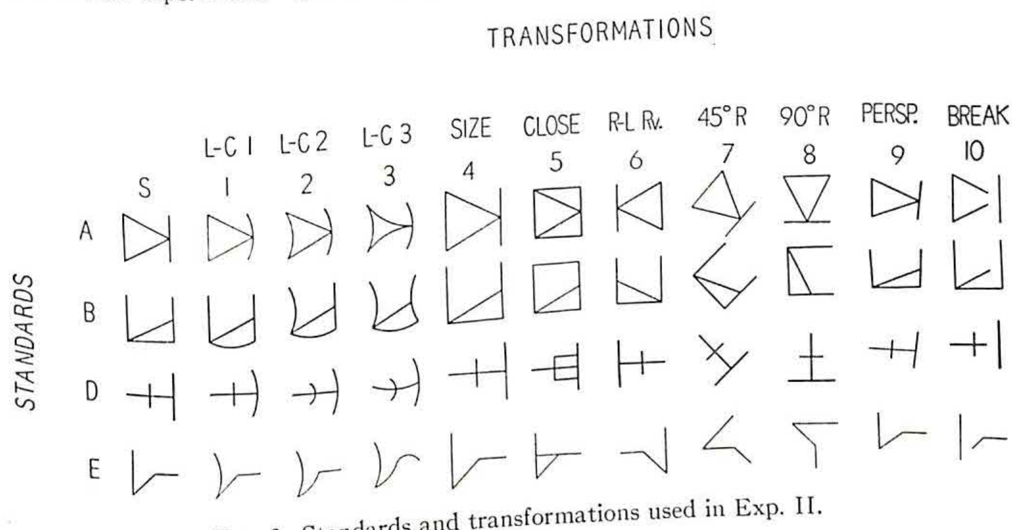


Fig. 2. Standards and transformations used in Exp. II.

two copies of the standard in random order. When S completed a trial, the procedure was repeated in a different random order until S reached the criterion of learning.[2] This criterion consisted of *either* a perfect trial *or* a single confusion error for either the size or perspective transformation (whichever the particular S was given). This weaker second condition was accepted as a criterion condition because size and perspective transformations proved in pilot work to be impossible for many *S*s to distinguish from standards long after they were able to distinguish the other transformations from standards.

Preliminary practice prior to training consisted of practice with visual letters followed by both visual and blindfolded practice with letters in the form of tactual stimuli.

Transfer.—The same three transfer conditions as before were used. Group C received a different standard and five different transformations of the standard from the ones used in training. Group EI received the same standard as in training, but five new transformations of this standard. Group EII received a new standard, but the same five transformations of this standard as those with which they had dealt in training.

Eight subgroups were used in the training condition representing training with each of the four standards and the two groups of transformations. One transformation group consisted of line to curve transformations, the size transformation, and the close. The other group consisted of the reversals, rotations, perspective transformation, and the break.

[2] The training task proved too easy for some first graders. Since this was a learning experiment, *S*s who performed the training task successfully on the initial trial were discarded from the experiment. There were also a few *S*s who were never able to learn the training task and who were also discarded. In all, about 100 *S*s were run through the training procedure. The final N represents those *S*s who did not perform at criterion level on the first trial and who were eventually able to do so after several training trials.

Results

Training

Differences between the groups in number of confusion errors made on the first trial and in number of trials to criterion were analyzed with t tests. None of the differences approached significance.

Transfer

Confusion errors made by the subgroups and for the three groups with subgroups combined are shown in Table 2. The analysis of variance indicated a significant effect of transfer groups at less than the .001 probability level, $F\ (2, 48) = 10.63$. The Transformations $\times$ Standards interaction and the triple interaction were also significant at less than the .05 level of probability. A comparison of the total errors made by the three groups indicated that Groups EI and EII were not different from each other but both were superior in performance to Group C.

TABLE 2

CONFUSION ERRORS IN TRANSFER TRIAL FOR THE 24 SUBGROUPS

Group	Transformations								Total
	T 12345				T 6789 & 10				
	Standards								
	SA	SB	SD	SE	SA	SB	SD	SE	
EI	3	8	2	2	4	1	1	3	24
EII	3	2	3	3	3	3	5	3	25
C	9	9	9	2	2	5	5	7	48

Note.—Column headings indicate combination of forms used in the transfer trial. $n = 3$ in each subgroup.

The Transformations × Standards interaction effect seemed to be due to a difference between errors made to Standard E and errors made to the other standards (cf. Fig. 2). For Standards A, B, and D more errors occurred with Transformations 1–5 than with Transformations 6–10. This pattern is reversed with Standard E which probably accounted for the fact that the effect of transformations did not reach significance.

The three-way interaction is difficult to interpret meaningfully. Probably its effect was accounted for by the fact that there were only three *S*s in a cell, and one atypical *S* in a given cell could account for the scores in that cell deviating from the overall pattern.

Discussion

Clearly the results of this experiment were different from those of the previous experiment. In that experiment, the group which had opportunity to use in transfer what it had learned about the *distinctive features* of the forms was superior to both the other groups in transfer trial performance. In the present experiment, the comparable group of *S*s was no better in transfer trial performance than the group which had opportunity to use prototypes or memory models of the forms, though both groups were better than the control group.

One interpretation of these results is that the process of improvement in visual discrimination is different from the process of improvement in tactual discrimination and that schema construction and detection of distinctive features may serve equally useful functions in tactual discrimination.

Another interpretation is related to differences in the task required of *S*. Because *S* in the tactual experiment explored the forms with only one hand, he made successive comparisons. This task involved memory and perhaps required him to form a memory model of the standard even if his task in learning was to detect distinctive features. In the visual experiment, however, *S* could look back and forth between the comparison forms without having to remember how they looked in order to discover the differences between them.

If this second interpretation is correct, making the tactual comparison simultaneous should then result in the reappearance of superior performance by Group EII. The third experiment was conducted to test this hypothesis.

Experiment III

The major difference between this experiment and the previous one is in the nature of *S*'s task. There were minor differences in number of *S*s, material, and subgroups as noted below.

Method

Subjects.—The *S*s were 60 first graders.

Materials.—The forms were Standards A and D and the 10 transformations of each.

Procedure.—The procedure was the same as in the previous experiment except that throughout this experiment, *S* explored the two comparison forms simultaneously, one with each hand.[3]

There were four subgroups representing training with each of the two standards and the two groups of five transformations each.

Results

Training

None of the differences between the groups in either number of trials or errors in the first trial approached significance. Hence the groups could

[3] Some *S*s could not complete the training task successfully and were discarded. Their inability to perform successfully appeared to be a function of an inability to coordinate both hands in exploring the forms. There were also a few *S*s who performed the task at criterion level on the first trial, and were discarded. About 80 *S*s began the training procedure and the final *N*, as in the previous experiment, represents those *S*s who did not perform at criterion level on the first training trial and who eventually learned the task.

TABLE 3

Confusion Errors in Transfer Trial for the 12 Subgroups

Group	Transformations				Total
	T 12345		T 6789 & 10		
	Standards				
	SA	SD	SA	SD	
EI	10	11	6	4	31
EII	3	3	3	0	9
C	10	10	5	7	32

Note.—Column headings indicate combination of forms used in transfer trial. $n = 5$ in each subgroup.

be considered equivalent with respect to their ability to handle the forms.

Transfer

Transfer trial errors for each subgroup and for subgroups combined are shown in Table 3. The analysis of variance performed on the data showed the effect of transfer groups to be significant with a probability of less than .001 that the effect was due to chance, $F\,(2, 48) = 8.62$. The effect of transformations was also significant with a probability of less than .01 that the effect was due to chance. Neither the effect of standards nor any interaction was significant.

A comparison of the total errors made by the three groups showed that Group EII, the group having new forms but familiar transformations in the transfer task, was superior to the other two groups in performance on this task. Group EI, the group having familiar forms but new transformations in the transfer task was not different from the control group.

The effect of transformations is due to the fact that more errors occurred with Transformations 1–5 than with Transformations 6–10. Except for one standard (E) this was also true in Exp. II. Apparently line to curve transformations are, in general, more difficult to discriminate tactually than rotations and reversals.

Discussion

These results support the hypothesis that under conditions of simultaneous comparison, the group having opportunity to use in transfer what they had learned about the distinctive features of the forms would show superior performance. Not only was this group (EII) superior to the other two groups, but the other experimental group showed no better performance than the control group. In this experiment, the construction of schemata, if such a process occurred at all, showed no effect in the transfer task.

General Discussion

A consideration of the tasks involved in these experiments may make the three different patterns of results meaningful. The third experiment involved a task of simultaneous comparison. The results suggested that *S*s had, in training, learned the distinctive features of the forms since the superior group had no opportunity to construct schemata of the forms used in the transfer task. Those *S*s who *could* use schemata in the transfer task performed no better than the control group.

The second experiment involved a task of *successive* comparison. These *S*s apparently both constructed schemata of the forms *and* learned distinctive features since groups who could use either distinctive features or prototypes showed similar amounts of transfer relative to the control group.

The first experiment involved a task which might be considered to lie between the tasks of the two tactual experiments in terms of the nature of the comparison. None of the *S*s in this experiment had to explore one form thoroughly before exploring the comparison form as did *S*s in the second experiment. On the other

hand, *S*s in this first experiment probably did not receive information from both the standard and comparison forms simultaneously as did *S*s in the third experiment. Most likely, *S*s in this first experiment quickly looked back and forth several times between the standard and comparison form in order to make a judgment about them. The *S*s who could use in transfer what they had learned about distinctive features showed the best transfer task performance. Those *S*s who could use schemata also showed transfer but significantly less than the other experimental group.

In terms of the tasks involved in these experiments, one might interpret the results as suggesting that the detection of distinctive features will always facilitate improvement in discrimination but that under conditions of successive comparison, schema construction will *independently* facilitate such improvement.

A more parsimonious interpretation is that the detection of distinctive features may be the *basis* for improvement in discrimination. When such detection is dependent on memory because of the nature of the task (e.g., in Exp. II and to a lesser degree in Exp. I), some schema learning does occur. When no memory requirement is imposed by the task (e.g., Exp. III), schema learning does not occur. In short, detection of distinctive features may be the necessary and sufficient condition for improvement in discrimination. Schema learning may or may not occur depending on the experimental conditions. When it does occur, its function is to make possible the comparison and search for differences, i.e., to make possible the detection of distinctive features.

The data of the present experiments are consistent with this interpretation. In no case did an EI group perform better than an EII group. Furthermore, the EI groups showed better performance than the control group only to the extent that memory was involved in the task of comparison.

Further research is necessary to establish the validity of this interpretation. A direct test might be made by determining whether, under conditions of successive comparison, *S*s in fact have learned a prototype or memory model of the standard forms. Can they identify the given standards from a group of unrelated forms, or can they reproduce the standards better than *S*s who have operated under conditions of simultaneous comparison?

REFERENCES

Bruner, J. S. Neural mechanisms in perception. *Psychol. Rev.*, 1957, **64**, 340–358. (a)

Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152. (b)

Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. A developmental study of the discrimination of letter-like forms. *J. comp. physiol. Psychol.*, 1962, **55**, 897–906.

Gibson, J. J., & Gibson, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.

Jakobson, R., & Halle, M. *Fundamentals of language.* The Hague, Netherlands: Mouton, 1956.

Solley, C. M., & Murphy, G. *Development of the perceptual world.* New York: Basic Books, 1960.

Vernon, M. D. *A further study of visual perception.* Cambridge, England: Cambridge Univer. Press, 1952.

Vernon, M. D. The functions of schemata in perceiving. *Psychol. Rev.*, 1955, **62**, 180–192.

(Received February 5, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 340–342

CUE AND SECONDARY REINFORCEMENT EFFECTS WITH CHILDREN[1]

JOSEPH B. SIDOWSKI, NORMAN KASS, AND HELEN WILSON

San Diego State College

Groups of children served in a lever-pulling experiment in order to determine the secondary reinforcement, cue, and combined secondary reinforcement/cue effects of a light stimulus. On each conditioning trial the light was differentially associated with the presentation of a primary reinforcement. After conditioning, Ss underwent a period of extinction during which time a lever pull resulted in a light or a no-light condition. An evaluation of the number of responses during extinction showed a significant difference between the secondary reinforcing and the cue effects of the light stimulus. The discussion covered a need for adequate experimental controls in studies attempting to show a "real" secondary reinforcement effect.

The purpose of the present study was to investigate the cue and secondary reinforcement (S^r) effects of a stimulus in an experiment utilizing children as *S*s. Cue properties of a stimulus were indicated when there was a tendency for a response to occur following the presentation of a stimulus. The stimulus showed reinforcing properties when there was an increase in the strength of some response that preceded the onset of the stimulus.

Although the cue and S^r effects of a stimulus have been discussed for a number of years, relatively few studies have been directed toward experimentally differentiating these effects (Bugelski, 1956; Elam, Tyler, & Bitterman, 1954; Melching, 1954; Wyckoff, Sidowski, & Chambliss, 1958). The lack of research in this area has been especially prevalent in human learning where the emphasis has been upon substantiating the existence of S^r and upon defining some of the variables influencing its strength (Hubbard, 1951; Myers, 1958; Myers & Myers, 1963).

Bugelski (1956) criticized conventional animal Skinner box tests of S^r on the grounds that the assumed secondary reinforcing signal may be evoking or facilitating the next lever response, rather than acting as a reinforcement for the preceding one. Wyckoff, Sidowski, and Chambliss (1958) found no indication of secondary reinforcing effects in either of two rat studies evaluating the relationship between the S^r and cue effects of a stimulus. The results of the above study were discussed with regard to the failure of other *E*s to control for cue error and suggest implied S^r effects to be due to the cuing of the stimulus. Recent reviews discuss the concept in detail and agree on a need for some clarification of the process (Myers, 1958; Wyckoff, 1959).

The present study was designed to differentiate the potential cue and S^r effects of a light stimulus during the conditioning of a lever-pulling response and to measure these effects

[1] This study was supported by funds from National Institute of Mental Health Grant MH 06806-01. The authors wish to express their appreciation for the cooperation of Burton Tiffany, Superintendent of Schools, James Spencer, Principal, and the teachers of Lauderbach Elementary School at Chula Vista; William Link, Principal, and the teachers of Kellog Elementary School; and Robert Nardelli, San Diego State College Laboratory School.

during extinction. If S^r effects exist and/or are unconfounded by cue effects, we would expect greater resistance to extinction for S^r in a design which allows for a differentiation of the two phenomena.

Method

Subjects.—The *S*s were 140 kindergarten children between the ages of 5.3 and 6.3 yr. Equal numbers of males and females were randomly assigned to each experimental condition.

Apparatus.—The *S*'s panel consisted of a simulated slot machine. A handle, mounted on the right side, could be pulled down to dispense pennies into a cup at the bottom of the front panel. A plastic hemisphere (dome) protruded from the center of the machine. Two 7½-w. red lamps were mounted directly behind a frosted upper portion of the dome. The lower half of the dome was clear and displayed a large number of pennies throughout the experiment. The presentation of light stimuli, dispensing of pennies, and recording of responses were controlled by electronic programing and recording equipment located in a room to the rear of the experimental room.

Two large pegboards displayed approximately 50 prizes to each *S*. The prizes consisted of toy watches, rings, parasols, fans, and other toy items of interest to children.

Procedure.—The *E* took each child from his classroom to an experimental room located in a quiet section of the building. The *S* was told that he was going to play the "penny machine game" and stood in front of the apparatus. The *E* pointed to the pile of pennies located in the dome and told *S* that he was to try to obtain as many as he possibly could. The *S* was then shown the entire display of prizes and told that the toys could be bought with the pennies. Each *S* was required to indicate the prize that he would like to obtain, following which he was given these instructions:

> Now I will show you how to play the game. When you pull the handle, the pennies come out here. You can play the game as long as you like. Tell me when you want to stop. Remember, the more pennies that you get out of the machine, the better chance you have of obtaining the prize that you want.

The *S*s were assigned to one of four experimental conditions: primary reinforcement (P^r), secondary reinforcement (S^r), Cue, and combined S^r/Cue effects. The P^r group received a penny each time the lever was pulled to the "down" position. The S^r group had a red light associated with the presentation of the penny during conditioning. The light was off until *S* pulled the handle to the down position; it automatically illuminated .5 sec. before the penny was dispensed and remained lit until the penny appeared in the cup, 1 sec. In the Cue group, the red light was presented at the beginning of each conditioning trial, i.e., while the handle was still in an "up" position. The light remained illuminated until *S* pulled the handle down, and was turned off *before* the penny appeared in the well. In the S^r/Cue group, the red light was presented while the handle was in the "up" position and remained illuminated until the lever was pulled down and the penny appeared in the well. It then went "off" until the next trial. Thus, the effects of reinforcement and cue were confounded.

Immediately after the 10 conditioning trials, *S*s in the S^r, Cue, and S^r/Cue conditions were subjected to one of two extinction conditions, Light or No Light. The red light was illuminated following the responses of the Light-only extinction groups; responses made by the No-Light groups were followed by neither pennies nor light. The *S*s in the P^r group served as an additional control and continued to receive pennies for lever pulling during the extinction interval. The design, therefore, had seven subgroups.

Twenty *S*s (10 male and 10 female) served in each of the seven extinction conditions. The temporal relationship of the light to the response was the same during extinction as it was during conditioning, e.g., the light came on for the S^r group only when the lever was in the "down" position. Each *S* in all groups played until he indicated a desire to stop, or until he completed 100 extinction trials.

Comments and questions from *S* were ignored by *E*. If *S* persisted in asking questions, *E* merely said: "You can play the game as long as you like. Tell me when you want to stop playing." This was generally sufficient. At the end of the session *S* was told that he would receive his toy after everyone in his room had played the game.

Results and Discussion

The raw scores of number of responses during extinction were transformed into log values to reduce heterogeneity of variance. Table 1 presents the mean log number of responses during extinction for each of

TABLE 1

Mean Log Number of Responses during Extinction

	S^r	Cue	S^r/Cue	P^r
Light	1.19	.80	.94	1.61
No light	.98	.76	.75	

the conditions. An analysis of variance on the log values indicated a significant difference between groups, $F(6, 133) = 8.15$, $p < .01$. Subsequent Duncan's test for ranked means showed the P^r continuous penny group producing significantly more lever-pulling responses than any extinction condition, $p < .01$. The S^r/Light subgroup showed greater resistance to extinction than the Cue/Light, Cue/No-Light, or S^r/Cue/No-Light conditions, indicating strong support for a reliable S^r effect. No other statistically significant differences were found.

For the S^r subgroups, classical secondary reinforcement literature would lead us to expect more resistance to extinction for the Light condition. The nonsignificant difference between Light and No Light is difficult to explain. An inspection of the original raw score data showed a definite separation of the subgroups in the number of responses above and below the median, 8. A median test (Siegel, 1956) showed these differences to be significant, $p < .01$, thus supplying additional support for the secondary reinforcement hypothesis.

Generally, the results of the present study showed that: (*a*) the light served as an S^r when it was associated directly with the presentation of the P^r, (*b*) if the light was introduced only while the lever was being pulled down, i.e., slightly prior to but not during the P^r interval, the number of responses during extinction was significantly fewer than resulted from the simultaneous presentation of the P^r and light, and (*c*) a combination of S^r and cue (a common practice in many S^r experiments) sustained lever-pulling behavior at a level somewhere between the pure S^r and pure cue conditions. Unfortunately, published studies often fail to report the exact specifications of the temporal sequence of the light (or sound), although this appears to be an important determiner of cue and S^r effects. If an *E* introduces a stimulus and response strength increases or remains high, then the experimental procedure must show that this increase was not due to a cue effect before citing the result as evidence for secondary reinforcement.

REFERENCES

Bugelski, R. *The psychology of learning.* New York: Holt, 1956.

Elam, C. B., Tyler, D. W., & Bitterman, M. E. A further study of secondary reinforcement and the discrimination hypothesis. *J. comp. physiol. Psychol.*, 1954, **47**, 381–384.

Hubbard, W. R. Secondary reinforcement of a simple discrimination in human beings. *J. exp. Psychol.*, 1951, **41**, 233–241.

Melching, W. H. The acquired reward value of an intermittently presented neutral stimulus. *J. comp. physiol. Psychol.*, 1954, **47**, 370–373.

Myers, J. L. Secondary reinforcement: A review of recent experimentation. *Psychol. Bull.*, 1958, **55**, 284–301.

Myers, J. L., & Myers, N. A. Secondary reinforcement in children as a function of conditioning associations, extinction percentages, and stimulus types. *J. exp. Psychol.*, 1963, **65**, 455–459.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Wyckoff, L. B. Toward a quantitative theory of secondary reinforcement. *Psychol. Rev.*, 1959, **66**, 68–78.

Wyckoff, L. B., Sidowski, J. B., & Chambliss, D. J. An experimental study of the relationship between secondary reinforcing and cue effects of a stimulus. *J. comp. physiol. Psychol.*, 1958, **51**, 103–109.

(Received February 3, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 343–348

TEMPORAL DETERMINANTS OF A KINESTHETIC AFTEREFFECT

G. SINGER AND R. H. DAY

University of Sydney

The development of a kinesthetic aftereffect as a function of stimulation times of 15, 30, 45, 60, 75, 90, 105, and 120 sec. and its dissipation after 30 and 90 sec. stimulation have been investigated in 2 experiments. The task required kinesthetic judgment of the horizontal following controlled rhythmic stimulation of the extended hand across a slanted bar. In Experiment I the aftereffect was shown to increase with stimulation time. In Experiment II the size of the aftereffect was similar immediately following 30 and 90 sec. stimulation but the rate of dissipation was greater for the shorter than for the longer stimulation. Since with the method of adjustment dissipation is rapid during the adjustment period, the development function of Experiment I is interpreted as representing a joint function of stimulation time and differential dissipation rates.

Following appropriate stimulation of those sensory systems capable of spatial discrimination, changes occur in judgments of size, shape, orientation, or position. These spatial aftereffects are now known to occur in vision, hearing, kinesthesis, and in the vestibular and tactile systems. In vision they have been extensively studied as the figural aftereffect.

From the viewpoint of a general theory it is of interest to establish the functional relationships between these effects and intensity, spatial, and temporal variables. The temporal determinants of the spatial aftereffect from kinesthetic stimulation have been investigated by Bakan, Myers, and Schoonard (1962), Bourne and Beier (1961), and Carlson (1963), all of whom used the now traditional task involving judgments of the width of a bar before and after stimulation by one of different width.

In an earlier study (Day & Singer, 1964) consistently large (4–5°) and reliable kinesthetic aftereffects were obtained in a task requiring manual judgments of the horizontal after motion of the extended arm and hand across a slanted bar. This task was suggested by the investigations of Gibson (1933) and actually used by Köhler and Dinnerstein (1947) who reported no quantitative data. So far little quantitative data are available on the determinants of this particular form of the kinesthetic aftereffect. It is of interest, however, to establish the relationship between this easily measurable aftereffect and those variables which are known to affect other kinesthetic aftereffects. The aim of the experiments reported here was to establish the relationship between the magnitude of this kinesthetic aftereffect and time of stimulation on the one hand, and its dissipation over time on the other.

EXPERIMENT I

The first experiment was concerned with the relationship between the size of the kinesthetic aftereffect and duration of stimulation, i.e., with the development of the effect with stimulation. In earlier experiments using the width judgment task (Bakan et al., 1962; Bourne & Beier, 1961; Carl-

son, 1963) three to five stimulation periods were used. Here eight durations ranging from 15 sec. to 120 sec. were used in order to obtain as detailed a function as possible.

Method

Apparatus.—The apparatus was adapted from similar equipment used in an earlier investigation (Day & Singer, 1964). It consisted of a metal lumber frame 46 in. high, 24 in. wide and deep in which was mounted a centrally pivoted 12-in. long bar in *S*'s frontal plane. The angle of bar could be adjusted by *S* or by *E* using suitably placed rotary controls operating through a 12:1 reduction gear. The angle of the bar could be read to the nearest .25° from an 8-in. diameter protractor scale at the rear of the frame.

Subjects.—Ten high school students (two boys, eight girls) whose ages ranged from 13 to 17 yr. (median age 14 yr.) acted as *S*s. No *S* was familiar with either the task or with kinesthetic aftereffects.

Experimental procedure and design.—A trial consisted of a pretest, a stimulation period, and a posttest. In the pre- and posttests, *S*, wearing opaque goggles, adjusted the bar with the left-hand control so that it seemed horizontal to the right hand moving at a preferred rate across its upper edge. To minimize the time taken to make an adjustment the bar was always set by *E* at the true horizontal prior to each pre- and posttest. Stimulation consisted of rhythmic side-to-side motion of the right hand across the upper edge of the bar slanted at 15° from the true horizontal. Hand motion was made in time with a metronome set at 84 beats per min. The difference between pre- and posttest adjustments in degrees was the measure of the aftereffect. On completion of stimulation *S* removed his hands from the apparatus while *E* quickly adjusted the bar to the physical horizontal.

Following five practice adjustments and appropriate instructions each *S* underwent two trials for each of eight stimulation periods of 15, 30, 45, 60, 75, 90, 105, and 120 sec. duration. Earlier observations had shown that motion of the extended hand across the slanted bar could not be sustained without discomfort for more than 2 min. In one trial of each pair the bar was slanted upward on *S*'s right during stimulation and downward in the other, the order of presentation of these orientations alternating from *S* to *S*. The eight pairs of trials were presented in a different random order for each *S* with a 3-min. rest period between each trial. The 3-min. interval was chosen since preliminary observations had shown that the effect had almost completely disappeared after this period. The possible occurrence of a cumulative aftereffect with repeated trials seems doubtful since the investigation of Heinemann (1961). The time taken by *S*s to complete the posttest adjustment was recorded and found to have a mean of 17 sec. Each session lasted about 90 min.

Results

The mean differences between pre- and posttests each based on two trials per *S* are shown for the eight stimulation periods in Table 1 along with their *SD*s.

The relation between size of aftereffect and stimulation time may be described as exponential. The function $y = A + B(1 - C^{x-1})/(1 - C)$ describes a family of monotonic curves ranging from linear (in the limit) to curvilinear. A computer program[1] exists for finding by iterative means estimates of the parameters A, B, and C for that one of the family or curves which is the best fitting line in the least-squares sense. Thus the program considers the linear fit as well as a family of curvilinear fits.

In the above expression, x is an ordinal number listing equally spaced values of stimulation time 1, 2 . . . j . . . t, y is the value of the aftereffect at the jth value of stimulation time, A is the value of the aftereffect for the first stimulation time, B is the change in the aftereffect between the first and second stimulation times (15 and 30 sec.), and C is the constant ratio of successive changes.

The least-squares estimates of A, B, and C were found for each individual

[1] The FORTRAN computer program for curve fitting was developed by J. P. Sutcliffe whose assistance in the treatment of these data and in the preparation of this report is gratefully acknowledged.

TABLE 1

MEANS AND *SD*s OF DIFFERENCES BETWEEN PRETESTS AND POSTTESTS FOR THE TWO EXPERIMENTS

Exp. I

	Stimulation Times (Sec.)							
	15	30	45	60	75	90	105	120
$\bar{X}$	1.80	1.76	2.22	2.60	2.72	2.52	3.04	3.11
SD	1.29	1.58	1.47	1.71	1.44	2.22	2.14	2.15

Exp. II

Cond.		Time after Cessation of Stimulation (Sec.)									
		0	15	30	45	60	75	90	105	120	180
30 sec.	$\bar{X}$	+2.69	+1.60	+1.01	+0.59	+0.19	+0.18	+0.83	−0.02	+0.09	+0.30
	SD	1.29	1.38	1.12	0.88	0.98	1.12	1.08	0.84	1.33	0.68
90 sec.	$\bar{X}$	+3.39	+3.24	+2.73	+1.30	+1.78	+1.11	+0.92	+0.99	+0.64	−0.09
	SD	2.00	2.03	2.12	1.59	1.88	1.86	1.27	1.32	1.35	1.00

S and for the mean data of the group. In the first case (mean parameter curve) the general function for the average values of *A*, *B*, and *C* is specified and, in the second case (mean performance curve) the curve is based on the *A*, *B*, and *C* values established from the group means.

The mean performance curve is shown in Fig. 1 as a continuous line. This plot suggests that after 2 min. stimulation the maximum aftereffect had not been reached. The mean parameter curve is shown as a dotted line in Fig. 1. Since not all *S*s exhibited the regular pattern of increment in the aftereffect with increasing stimulation time, this second plot differs from the mean performance plot. Both curves however, show an exponential increase in kinesthetic aftereffect as a function of stimulation time, the mean parameter curve tending toward an asymptote at an earlier stage than the mean performance fit.

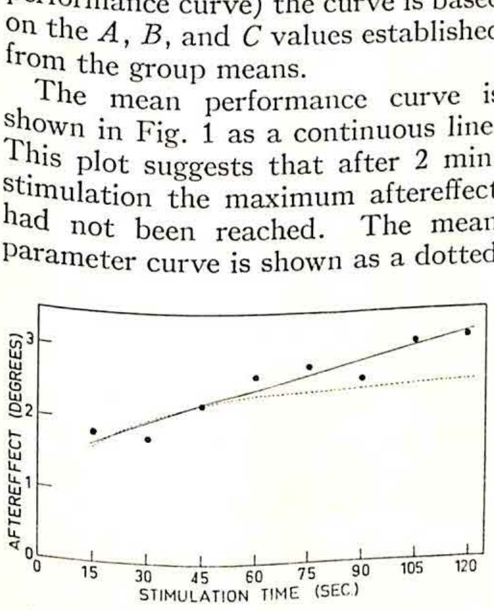


FIG. 1. Kinesthetic aftereffect as a function of stimulation time. (Continuous line: mean performance curve; dotted line: mean parameter curve.)

EXPERIMENT II

There is good reason for supposing that the development function in Exp. I represents an interaction between stimulation time and different dissipation rates of the aftereffect for different stimulation times. Oyama (1953) in an extension of earlier experiments by Hammer (1949) showed that the dissipation rate of a visual-spatial aftereffect was relatively fast for short stimulation periods and slower for the longer stimulation durations. Since in Exp. I there was a mean time lapse of 17 sec. for the posttest, it is

possible that for the shorter stimulation periods especially, much of the effect would have dissipated by the completion of the adjustment. Less dissipation would be expected to occur for the longer stimulation periods.

The aim of Exp. II was to establish the dissipation functions for the kinesthetic aftereffect for two stimulation times: 30 sec. and 90 sec. If these functions varied in rate then it would be reasonable to assume that the curves of Fig. 1 are joint functions of development and dissipation, the latter being an artifact of the adjustment method.

Method

Apparatus.—The same apparatus as in Exp. I was used.

Subjects.—There were two groups of 12 adult male and female *S*s all of whom were either graduate or senior students in the Department of Psychology. The *S*s varied in their knowledge of spatial aftereffects but they had not previously performed the particular task.

Experimental procedure and design.—Apart from stimulation times and the intervals elapsing between cessation of stimulation and posttest, all conditions were the same as in Exp. I including the alternation of directions of tilts and the 3-min. interval between trials. After cessation of stimulation *S* rested with hands on knees and, at "a" signal from *E*, the hands were replaced and the posttest adjustments made. In the 0-sec. condition the procedure was the same as in Exp. I.

Each *S* underwent 10 randomized pairs of trials in two blocks of five with at least 1 day between blocks. The 10 intervals between cessation of stimulation and initiation of the posttest were 0, 15, 30, 45, 60, 75, 90, 105, 120, and 180 sec. For half the *S*s in each group this order was reversed. One group of 12 *S*s underwent 30 sec. stimulation and the other group 90 sec. A record was kept of the times taken to complete the posttest adjustment. These times had a mean of 12 sec.

Results

In Table 1 are shown the mean differences and *SD*s between pre- and posttests for the 10 intervals between cessation of stimulation and commencement of posttest under the two stimulation conditions (30 sec. and 90 sec.). Following the procedure adopted in Exp. I, two curves have been fitted to these data: one based on the mean performance and the other derived from the mean of the individual parameters. The mean performance curves for the two conditions are shown as continuous lines in Fig. 2 and mean parameter curves as dotted.

Since the mean time taken for *E* to reset the bar and for *S* to make the adjustment was 12 sec., this time has been added to each of the intervals so that 0 sec. becomes 12 sec., 15 sec. becomes 27 sec., etc. It should be noted, however, that since the program used for fitting the curves requires equal intervals between values of the independent variable (i.e., between delays from cessation of stimulation to posttest) the curve has not been fitted to the 180-sec. data point (i.e., 192-sec. plotted point in Fig. 2).

Whereas there is little difference between the two curves for the 90-sec. stimulation condition, for the 30-sec. condition the mean parameter curve

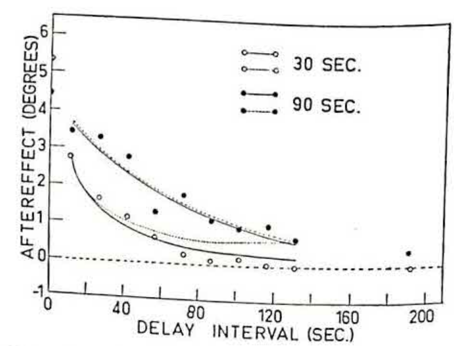


FIG. 2. Kinesthetic aftereffect as a function of time after cessation of stimulation periods of 30 sec. and 90 sec. (Continuous lines: mean performance curves; dotted lines: mean parameter curves. The computed points for intersection with the ordinate—i.e., 0-sec. delay—are also shown.)

tends to an asymptote at a higher value (about 0.6°) than the mean performance curve. This difference is probably due to one *S* who exhibited a dissipation function opposite in direction to the other 11 *S*s.

Since the original 0-sec. condition has become 12 sec. because of the time occupied by the posttest adjustment, it has been necessary to compute values for the true 0-sec. condition. These derivations are 5.3 for the 30-sec. stimulation condition and 4.4 for the 90-sec. condition, and both points are shown in Fig. 2 on the ordinate.

Discussion

If the data of Exp. I and II are considered together it is reasonable to conclude that the growth of the aftereffect as a function of stimulation time (Fig. 1) represents an interaction between the possible incremental effects of stimulation time and the dissipation effects occurring during the posttest adjustment. The method of adjustment necessarily involves a time lapse after stimulation. During that period the aftereffects dissipate at a rate determined by stimulation duration (Fig. 2). Thus on completion of the posttest, it would be expected that much of the effect would have dissipated during the 17-sec. adjustment period. Further, since the rate of dissipation can be assumed to be greater for the shorter periods of stimulation than after 17 sec., the aftereffect would be smaller for the shorter duration times than for the longer. It can be assumed, however, that using a psychophysical method with a shorter response latency (constant stimuli or limits) the development function would exhibit a steeper initial rise and achieve a maximum after a briefer stimulation period than that suggested by Fig. 1.

The results from Exp. II show that the initial sizes of the aftereffect soon after 30 sec. and 90 sec. stimulation are similar, differing by 0.9°. In point of fact, while the 30-sec. stimulation plot is for the most part below that of the 90-sec. in Fig. 2, its initial computed magnitude is greater. Both sets of data conform closely to an exponential decay function.

There is a striking resemblance between the functions found here for a kinesthetic aftereffect and those reported by Oyama (1953) for the visual aftereffect. In the latter experiments it was shown that following stimulation periods ranging from 1 sec. to 240 sec. the initial magnitudes of the effect were constant, but that dissipation was very rapid for the short periods and slow for the longer. It was also demonstrated by Oyama that, if the delay between stimulation and judgment was very short (1 sec.), the development function was rapid, achieving its maximum after about 25 sec. If the interval was controlled at 5 sec., however, the maximum effect did not occur until after 240 sec. stimulation.

The exponential functions found by Oyama (1953) for the development and dissipation of a visual-spatial aftereffect and those reported here for the kinesthetic effect suggest that these two aftereffects behave in a similar fashion as a function of stimulation time and time after stimulation.

In conclusion, a point of general methodological interest can be raised. The mean adjustment times for the younger group of *S*s in Exp. I was 17 sec., while that for the adult groups of Exp. II was 12 sec. The measured size of both visual and kinesthetic aftereffects using the method of adjustment is determined by both stimulation time and adjustment time. It is to be expected, therefore, that, if the adjustment times found here are a true indication, more of the effect will dissipate with younger than with older *S*s during the posttest. This expectation is critical for developmental and comparative studies of spatial aftereffects where differences in the size of the aftereffects are often claimed. For example, the smaller visual effect found by Prysiazniuk and Kelm (1963) for retarded adults after 40 sec. stimulation as compared with the effect manifested by normal *S*s could well be due to the different adjustment times for the two groups.

REFERENCES

BAKAN, P., MYERS, L. B., & SCHOONARD, J. Kinesthetic aftereffects and length of inspection period. *Amer. J. Psychol.*, 1962, **75**, 457–461.

BOURNE, L. E., & BEIER, E. G. Effect of duration of inspection upon kinesthetic figural aftereffects. *J. gen. Psychol.*, 1961, **65**, 163–169.

CARLSON, J. B. Effect of amount and distribution of inspection time and length of delay interval on kinesthetic aftereffect. *J. exp. Psychol.*, 1963, **66**, 377–382.

DAY, R. H., & SINGER, G. Spatial aftereffects within and between kinesthesis and vision. *J. exp. Psychol.*, 1964, **68**, 337–343.

GIBSON, J. J. Adaptation, aftereffect, and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.

HAMMER, E. R. Temporal factors in figural aftereffects. *Amer. J. Psychol.*, 1949, **62**, 337–354.

HEINEMANN, E. G. Figural aftereffects in kinesthesis: Effects of object width and repeated presentations. *J. exp. Psychol.*, 1961, **61**, 51–56.

KÖHLER, W., & DINNERSTEIN, D. Figural aftereffects in kinaesthesis. In Albert Michotte, *Miscellanea psychologica*. Louvain: Catholic University of Louvain, 1947. Pp. 196–220.

OYAMA, T. Experimental studies of figural aftereffects: I. Temporal factors. *Jap. J. Psychol.*, 1953, **23**, 239–254.

PRYSIAZNIUK, A. W., & KELM, H. Visual figural aftereffects in retarded adults. *J. abnorm. soc. Psychol.*, 1963, **67**, 505–509.

(Received February 3, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 349–356

EFFECTS OF SHOCK INTENSITY AND PLACEMENT ON THE LEARNING OF A FOOD-REINFORCED BRIGHTNESS DISCRIMINATION [1]

ELIZABETH R. CURLIN AND JOHN W. DONAHOE

University of Kentucky

The effects of shock intensity (.13 ma. or .30 ma.) and shock placement (S+ or S−) on the learning of a brightness discrimination were investigated. 30 albino rats served as Ss in a bar-pressing task. Low-intensity shock (.13 ma.) facilitated the rate of discrimination formation whether placed in S+ or S−. The omission of .13 ma. shock after the attainment of a stable discrimination left performance relatively unchanged. Moderate intensity shock (.30 ma.) facilitated discrimination performance when placed in S− and hindered performance when placed in S+. The omission of .30 ma. shock produced an immediate deterioration in the performance for those Ss shocked in S− and an improvement in performance for those Ss shocked in S+. It was concluded that low-intensity shock acts primarily as an additional stimulus to facilitate discrimination learning, whereas moderate intensity shock acts primarily as an aversive (motivational) stimulus to affect performance.

In the present experiment, a bar-pressing task was used to assess the effects of two intensities of shock on the learning of a food-reinforced brightness discrimination. A response in the positive stimulus (S+) produced shock for half the animals; a response in the negative stimulus (S−) produced shock for the remaining half. It was hypothesized that low-intensity shock serves predominantly as a discriminative stimulus (cue function) whereas high-intensity shock serves predominantly as an aversive stimulus (motivational function). This hypothesis is a particularization of Miller and Dollard's (1941) statement that, "The stronger the stimulus, the more drive function it possesses [p. 18]." At low intensities of shock, discrimination formation should be facilitated as the shock stimulus (whether it occurs in S+ or S−) increases the difference between the stimulus conditions prevailing during the reinforcement and the nonreinforcement of the bar-pressing response. At higher shock intensities, however, the aversive properties of shock should decrease the frequency of those responses which produce it and thereby facilitate discrimination formation only when placed in S− and hinder discrimination formation when placed in S+.

METHOD

Subjects.—Thirty male albino rats (Wistar strain, Harlan Industries), 80–90 days old, served as *S*s. One week before training all animals were reduced to 80% of normal body weight and were maintained at this level for the duration of the study. All animals were fed a measured daily ration 1 hr. after the completion of each experimental session. Throughout the experiment, *S*s were housed in individual living cages in a temperature-controlled room.

Apparatus.—Two Skinner boxes of identical construction were used. Each box was enclosed in a sound-deadened hull with a blower to furnish ventilation and a masking

[1] The research was supported by grants to the second author from the Faculty Research Fund, University of Kentucky, and from the Public Health Service (MH-07332-01). The authors are indebted to the Computing Center, University of Kentucky, for the use of their facilities in this project.

noise. A Gerbrands response bar was centrally located in one wall of each chamber with a food cup 1 in. to the left and a drinking tube 1 in. to the right of the bar. A pilot lamp with a ½-in. white jewel was mounted 2½ in. above the bar and served as the discriminative stimulus. The chamber was dimly illuminated at all times by a houselight. A constant current shock supply with a scrambler (Grason-Stadler E1064GS) was used to electrify the grid floors, walls, bar, food cup, and drinking tube. The experimental events were programed by means of standard relay and timing equipment and the responses were monitored by printing counters.

Procedure.—All *S*s received a pretraining period which consisted of three sessions of continuous reinforcement. During the first session, bar pressing was acquired through the method of approximations with 10–15 gratuitous pellets being delivered to each animal. The first session was terminated at the end of 1 hr. or after 120 reinforced responses, whichever occurred first. For the remaining two sessions of pretraining, a 45-mg. food pellet was placed in the food cup at the beginning of each session and bar pressing was continuously reinforced for 15 min. During pretraining, the houselight provided the sole source of illumination.

Discrimination training was then begun with continuous reinforcement scheduled when the pilot lamp was on (S+) and extinction scheduled when the lamp was off (S−). The various treatment conditions were formed by randomly subdividing *S*s into the following five equal groups: (*a*) no shock in either S+ or S− (a no-shock control group, Group C), (*b*) .13 ma. shock in S+ (.13 S+ group), (*c*) .13 ma. shock in S− (.13 S− group), (*d*) .30 ma. shock in S+ (.30 S+ group), and (*e*) .30 ma. shock in S− (.30 S− group). For the four experimental groups, the shock stimulus was .50 sec. in duration and was administered immediately after the depression of the bar. Of the six animals in each group, three were run in each of the two Skinner boxes. In summary, the experimental plan consisted of a 2 × 2 factorial design having shock intensity (.13 ma. or .30 ma.) and shock placement (S+ or S−) as variables of classification with the addition of a single no-shock control group.

Eight sessions of discrimination training under the appropriate shock conditions were given. A session was composed of 12 discrimination cycles, each cycle consisting of 120 sec. in S+ followed by 120 sec. in S−. For all experimental groups the shock intensity was gradually increased to the appropriate value. For the low-shock (.13 ma.) groups, shock intensity was .08 ma. for Cycles 1–6 of the first session and remained constant at .13 ma. thereafter. For the higher shock (.30 ma.) groups, the shock intensity was .08 ma. for Cycles 1–6 and .13 ma. for Cycles 7–12 of the first session, .13 ma. for Cycles 1–6 and .25 ma. for Cycles 7–12 of the second session, and .25 ma. for Cycles 1–6 of the third session with the intensity remaining constant at .30 ma. thereafter. The gradual increase in shock intensity avoided the complete suppression of bar pressing occasionally produced by an initial shock of .30 ma. and provided additional data on the effects of low-intensity shock in the early stages of discrimination learning. Thus the shock intensity was the same for the four experimental groups during the initial 18 cycles of training.

After eight sessions of discrimination training, all *S*s were given an additional two sessions in which shock was omitted. The changes observed in postshock performance were intended to provide information concerning the type of control exercised by shock stimulation during the first eight sessions.

Results

Performance was measured by two dependent variables: (*a*) the proportion of the total responses occurring to S+ and (*b*) the number of responses occurring to S+ and S−. The S+ proportion provided a summary statistic which indicated the level of discrimination directly while the number of S+ and S− responses is the primary data which determine the level of discrimination. Unless otherwise indicated, data from the first two discrimination cycles were excluded from the following analyses. This was necessitated as the .13-ma. and control groups consistently improved in performance within all training sessions whereas a "warm-up" effect of comparable magnitude did not occur with the .30-ma. groups. The first two cycles of each session were therefore eliminated to minimize the confounding of intergroup differences with unequal warm-up rates.

Effect of .13-ma. shock.—During the first 18 cycles of discrimination training when all experimental groups were

receiving a maximum shock intensity of .13 ma., the proportion of responses in S+ was greater for the experimental groups than for the control group, t (28) = 2.58, $p < .02$. Considering the performance of the .13-ma. groups across all training sessions, low-intensity shock facilitated the rate of discrimination formation whether placed in S+ or S− (see Fig. 1). The rate of change in the S+ proportion, as measured by least-squares estimates of the slope parameter,[2] indicated that the .13-ma. groups acquired the discrimination more rapidly than the control group, t (20) = 2.51, $p < .05$. The more rapid increase in the S+ proportion for the low-shock groups was primarily due to an abrupt decrease in the frequency of S− responding (see Fig. 2, Panel A). Whereas the frequency of S− responding decreased significantly between Sessions 1 and 2 in the .13 S+ and .13 S− groups, t (175) = 5.89, $p < .001$ and t (175) = 3.89, $p < .001$, respectively, the comparable difference was not reliable in Group C, t (175) = 1.54, $p < .20$. The number of responses in S+ tended to be greater than the control group in the .13 S− group and less in the .13 S+ group although neither difference was significant. The asymptotic level of discrimination, as measured by the S+ proportion during Session 8, was the same for Groups .13 S+, .13 S−, and C.

In the first no-shock session (Session 9), the S+ proportion remained unchanged for the low-shock and control groups (see Fig. 1). The fre-

[2] In order to secure a single measure of the rate of discrimination formation, the S+ proportion and session number underwent a reciprocal transformation appropriate to the linear form of the hyperbolic function, $Y = X/(a + bX)$. For all groups having positive slopes (.13 S+, .13 S−, .30 S−, and C), the median index of correlation was .923 for individual animal curves and .977 for group curves.

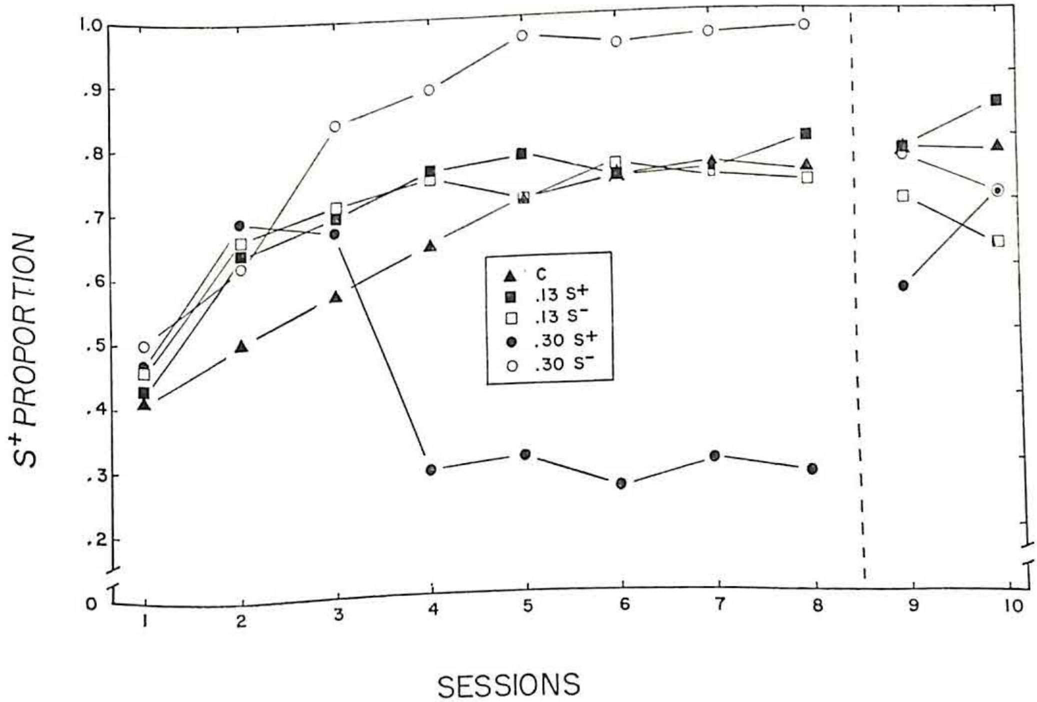


FIG. 1. Proportion of responses in the positive stimulus (S+) during shock sessions (1–8) and during no-shock sessions (9 and 10).

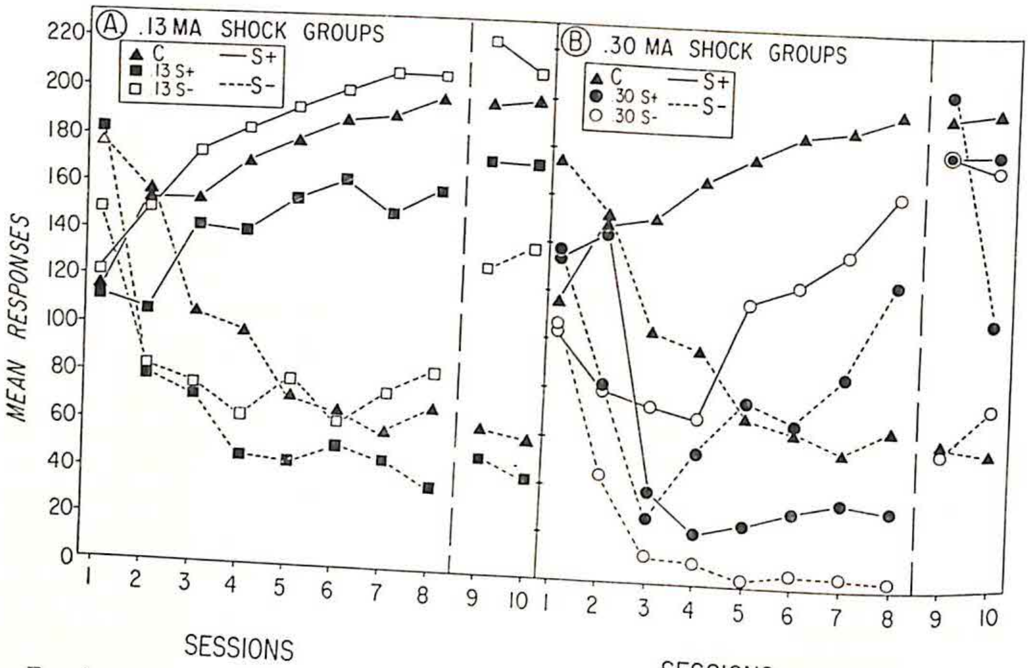


FIG. 2. Number of responses in the positive stimulus (S+) and negative stimulus (S−) during shock sessions (1–8) and during no-shock sessions (9 and 10). (The solid line represents S+ responding; the broken line represents S− responding.)

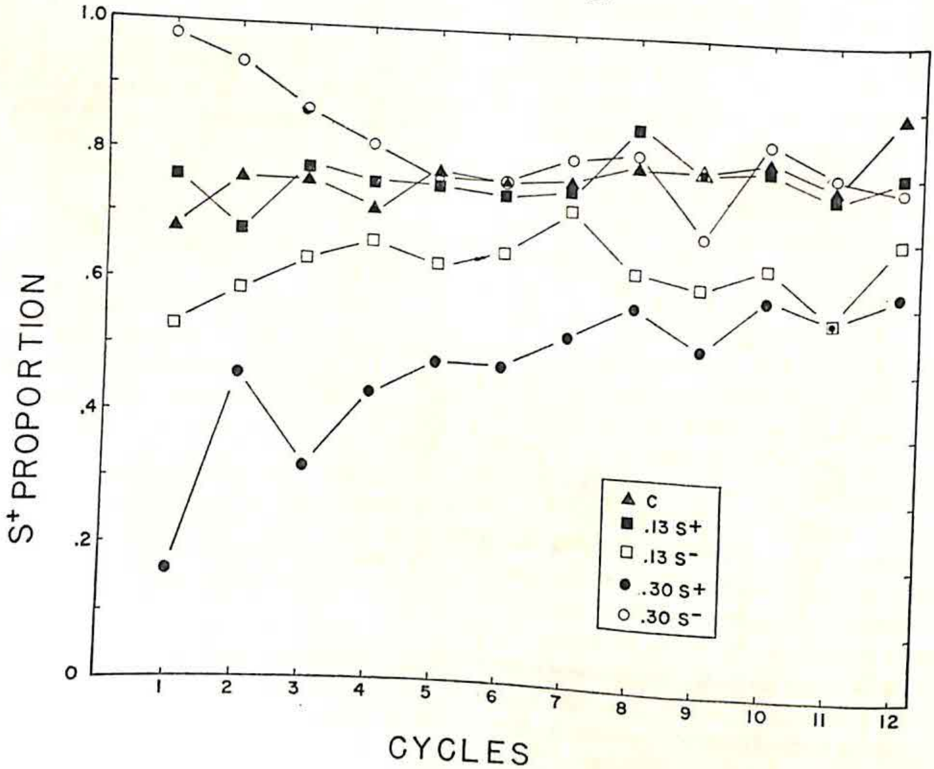


FIG. 3. Proportion of responses in the positive stimulus (S+) during the 12 discrimination cycles of the first no-shock session (9).

quency of responding was constant in Groups C and .13 S+, but increased in the .13 S− group particularly in S−, t (167) = 3.22, $p < .01$. Figure 3, in which the S+ proportion is plotted for each of the 12 discrimination cycles within Session 9, indicates that the .13 S+ and control groups performed at a similar level whereas the .13 S− group remained nonsignificantly below them. By the second no-shock session (Session 10), the S+ proportion had fallen significantly below the level of Session 8 for the .13 S− group, t (50) = 3.29, $p < .01$. As shown in Fig. 2, the decrease in the S+ proportion for the .13 S− group was due to an increase in S− responding. The .13 S+ group, by contrast, displayed an increase in the S+ proportion in the second postshock session, t (50) = 2.47, $p < .02$.

Effects of .30-ma. shock.—The .30 S+ group showed an initial increase in the S+ proportion followed by a decrease as the shock intensity reached .30 ma. (see Fig. 1). This trend, which was unlike that found in any other group, caused a significant cubic component, F (4, 171) = 7.28, $p < .001$, in the interaction of groups with sessions (see Table 1). The .30 S− group consistently produced a high proportion of responses in S+. The rate at which the discrimination developed in the .30 S− group did not differ, however, from the control group as measured by the slope estimates, t (20) = 1.33, $p < .20$. As shown in Fig. 2, Panel B, the initial effect of increasing the shock intensity in the .30-ma. groups was to decrease the frequency of responding in both S+ and S− regardless of the stimulus in which shock occurred. The decrease in response frequency between Sessions 1 and 3 was significant for the .30 S+ and .30 S− groups in both S+ and S− ($p < .05$ in every case).

TABLE 1

ANALYSIS OF VARIANCE OF THE PROPORTION OF RESPONSES IN S+ DURING DISCRIMINATION TRAINING

Source	df	F[a]
Between animals	29	
Control group vs. exp. groups	1	0.33
Shock level (L)	1	6.86*
Shock placement (P)	1	58.17***
L × P	1	63.44***
Animals within groups (A)	25	(0.036)
Within animals	206[b]	
Sessions (S)	7	18.50***
S × Groups	28	8.58***
linear	4	44.75***
quadratic	4	2.26
cubic	4	7.28***
residual	16	1.51
S × A[c]	171	(0.012)

[a] Values in parentheses are *MSs*.
[b] Due to failure of the printing counter, 4 df have been lost for estimation of the missing data.
[c] This source of variance is a pooled error term following the satisfaction of F tests for homogeneity of the various polynomial components (linear, quadratic, etc.).
* $p < .05$.
*** $p < .001$.

As the discrimination sessions continued, the S+ proportion for the .30 S+ group remained significantly below the control group, t (25) = 4.32, $p < .001$, and the S+ proportion for the .30 S− group remained significantly above the control group, t (25) = 2.08, $p < .05$. The difference in S+ proportion found between the .30-ma. groups, but not between the .13-ma. groups, gave rise to a significant interaction of shock placement with shock intensity, F (1, 25) = 63.44, $p < .001$ (see Table 1). The changes in response frequency were also different as a function of shock intensity. In the .30 S+ group, responding in S+ remained at a low level throughout training, but responding in S− increased until it was significantly *above* the S− response frequency of the control group at the eighth session, t (25) = 2.46, $p < .03$. Thus the S+ proportion of the .30 S+ group was depressed not only because

of a low response rate in S+, but also because of a relatively high response rate in S−. In the .30 S− group, the frequency of responding in S+ gradually increased after the initial decline but remained below the S+ frequency of Group C, t (25) = 3.67, $p < .01$. The S− frequency remained at a very low level throughout the shock sessions. A .30-ma. shock, therefore, produced different effects during discrimination training than a .13-ma. shock. Responding under .30-ma. shock was generally less frequent, F (1, 25) = 30.24, $p < .001$, and displayed a different trend in S+ and S− over sessions, F (28, 169) = 6.88, $p < .001$ (see Table 2).

The postshock behavior of the .30-ma. groups also differed from the .13-ma. groups. On the first postshock session, the S+ proportion of the .30 S+ group increased, t (25) = 7.28, $p < .001$ and the S+ proportion of the .30 S− group decreased, t (25) = 6.69, $p < .001$ (see Fig. 1). The abruptness of these changes in S+ proportion are shown in Fig. 3. The improvement of performance in the .30 S+ group and the deterioration of performance in the .30 S− group both became more pronounced on the second postshock day. The changes in responding which produced the alteration in the S+ proportion are shown in Fig. 2. The removal of shock in the .30 S+ group brought about a sharp increase in both S+ and S− responding, t (167) = 10.28, $p < .001$ and t (167) = 5.52, $p < .001$, respectively. On the second postshock day, S− responding declined, t (167) = 6.64, $p < .001$. In the .30 S− group, the removal of shock was followed by an increased response frequency in S−, t (167) = 3.78, $p < .001$. The increase in S− responding continued unabated during Session 10, t (167) = 5.03, $p < .001$.

TABLE 2

ANALYSIS OF VARIANCE OF THE FREQUENCY OF RESPONDING IN S+ AND S− DURING DISCRIMINATION TRAINING

Source	df	F[a]
Between animals	29	
Control group vs. exp. groups	1	34.21***
Shock level (L)	1	30.24***
Shock placement (P)	1	1.76
L × P	1	3.36
Animals within groups (A)	25	(7,417.)
Within animals	442[b]	
Sessions (S)	7	10.18***
S × Groups	28	2.75***
S × A	175	(1,239.)
S+ vs. S− discrimination (D)	1	90.34***
D × Groups	4	11.42***
D × A	25	(4,809.)
S × D	7	24.89***
S × D × Groups	28	6.88***
S × D × A	167	(921.)

[a] Values in parentheses are MSs.
[b] Due to failure of the printing counter, 8 df have been lost for estimation of the missing data.
*** $p < .001$.

DISCUSSION

A shock of .13-ma. intensity, whether placed in S+ or S−, produced a more rapid increase in the S+ proportion than found in the no-shock control group. In both .13-ma. groups, the increased rate of discrimination formation was due to an abrupt initial decline in the frequency of S− responding. This finding is interpreted as supporting the hypothesis that low-intensity shock stimulation facilitated the learning of the discrimination by increasing the difference between the stimulus conditions present during the reinforcement and nonreinforcement of the bar-pressing response. That is, shock served as another cue, in addition to light onset and the aftereffects of reinforcement, which became associated with bar pressing through differential reinforcement. Viewed in this way, the shock stimulus facilitated discrimination learning through the mechanism of habit summation previously documented in

the T maze (Eninger, 1952). The ability of shock per se to serve as a discriminative stimulus controlling the rate of responding has been demonstrated by Holz and Azrin (1961). Pigeons were trained to peck a key under conditions in which a 2-min. variable-interval reinforcement schedule alternated with extinction periods. Responding during the variable-interval schedule produced both food and shock. In extinction sessions, response-produced shock increased the frequency of pecking above the no-shock periods.

Even at an intensity of .13 ma., there was some tendency for shock to be aversive. In the .13 S+ group, S+ responding remained nonsignificantly below control levels during shock sessions and, in both .13-ma. groups, S+ and S− responding increased during postshock sessions. The postshock S+ proportion and response frequency remained more stable, however, than with .30-ma. shock. The most pronounced increase in postshock responding—that occurring in S− in the .13 S+ group—cannot be unequivocally interpreted as the removal of the shock stimulus could produce an increase in responding due to both the elimination of aversive stimulation and the institution of stimulus conditions more similar to those previously present in S+ (i.e., no shock). Of the two factors, the cue variable is probably the more important as S− responding was not depressed below control levels during the last four shock sessions in the .13 S− group. It should also be noted that both cue and motivational factors would favor an increase in response frequency only in the S− condition of the .13 S− group.

The effects of .30-ma. shock on performance were quite different from those obtained with .13 ma. The initial response to .30-ma. shock was a decreased frequency of responding to both S+ and S−. In the .30 S+ group, S+ responding remained at a low level throughout the shock sessions while S− responding gradually increased to a frequency above control levels. The reduction in S+ responding was clearly a result of the aversive nature of .30-ma. shock as responding rapidly rose to control levels on the first postshock day. The increase in S− responding in the .30 S+ group may be interpreted as an example of displacement (Miller, 1948, 1959). Thus S− responding, which was initially produced via generalization from S+ and pretraining, was thereafter maintained by fear reduction which occurred when bar pressing in S− was not followed by shock. In support of this interpretation, the frequency of S− responding in the .30 S+ group rose sharply in the first postshock session and then declined rapidly with continued exposure to no-shock conditions.

The .30 S− group attained a higher level of discrimination than the control group although the rate of approach to the asymptotic value of the S+ proportion was not more rapid. As further evidence for the predominance of aversion with .30-ma. shock, the superior performance of the .30 S− group was achieved through a marked depression in the frequency of S− responding. Shock performance of both .30-ma. groups was therefore consistent in indicating that response frequency covaried with the stimulus in which shock occurred rather than the stimulus in which food occurred. In postshock sessions, the frequency of S− responding increased rapidly in the .30 S− group. As with the postshock responding of the .13 S− group, a cue factor could have contributed to the increase in S− responding although a motivational factor provides the more likely interpretation in the .30 S− group because of the suppression of S− responding during shock sessions. The finding that punishment failed to permanently suppress nonextinguished S− responding is consistent with the work of Skinner (1938) and Estes (1944). In summary, .30-ma. shock in S− is most readily interpreted as having predominantly aversive properties which produced a greater proportion of responses in S+ through punishment of S− responding rather than through

facilitation of the learning of the brightness discrimination.

The finding that a low-intensity shock (.13 ma.) in either S+ or S− facilitated discrimination learning whereas a moderate intensity shock (.30 ma.) facilitated performance when placed in S− and hindered performance when placed in S+ is relevant to the Muenzinger-Wischner controversy concerning the "shock-right" effect. Although differences in procedure (Muenzinger and Wischner used modified T mazes) make detailed comparisons difficult, both sets of experiments suggest that shock intensity is a critical parameter which determines whether shock in S+ will facilitate or hinder discrimination performance. Muenzinger (1934; Muenzinger, Bernstone, & Richards, 1938) has found a facilitating effect of shock in S+ and has always employed a shock of low intensity (.10–.15 ma.).[3] Moreover, he has reported that facilitation is more pronounced if the aversive effects of shock are reduced through a pretraining adaptation procedure (Muenzinger, Brown, Crow, & Powloski, 1952). Wischner (1947), who found that shock in S+ hindered discrimination, used a shock of moderate intensity—.30 ma. In a later study (Wischner, Fowler, & Kushnick, 1963) in which shock intensity was varied, Wischner reported that discrimination was hindered progressively as shock intensity increased from .15 to .25 ma. The import of all investigations is therefore consistent in indicating that the facilitation of performance by shock in S+ is confined to low-intensity shock having minimal or no aversive properties. The fact that Muenzinger employed a correction method and Wischner a noncorrection method is also an important feature contributing to the differences in their results and has been fully discussed by both authors (Muenzinger & Powloski, 1951; Wischner, 1947).

[3] To facilitate a comparison of the various experiments, all shock intensities have been transformed to units of current (ma.) where necessary by assuming an animal resistance of 300,000 ohms.

REFERENCES

ENINGER, M. V. Habit summation in a selective learning problem. *J. comp. physiol. Psychol.*, 1952, **45**, 604–608.

ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57** (3, Whole No. 263).

HOLZ, W. C., & AZRIN, N. H. Discriminative properties of punishment. *J. exp. Anal. Behav.*, 1961, **4**, 225–232.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

MUENZINGER, K. F. Motivation in learning: I. Electric shock for correct responses in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 439–448.

MUENZINGER, K. F., BERNSTONE, A. H., & RICHARDS, L. Motivation in learning: VIII. Equal amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.

MUENZINGER, K. F., BROWN, W. C., CROW, W. J., & POWLOSKI, R. F. Motivation in learning: XI. An analysis of electric shock for correct responses into the avoidance and accelerating components. *J. exp. Psychol.*, 1952, **43**, 115–119.

MUENZINGER, K. F., & POWLOSKI, R. F. Motivation in learning: X. Comparison of electric shock for correct turns in a correct and a noncorrection situation. *J. exp. Psychol.*, 1951, **42**, 118–124.

SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

WISCHNER, G. J. The effect of punishment on discrimination learning in a non-correction situation. *J. exp. Psychol.*, 1947, **37**, 271–284.

WISCHNER, G. J., FOWLER, H., & KUSHNICK, S. A. Effect of strength of punishment for "correct" and "incorrect" responses on visual discrimination performance. *J. exp. Psychol.*, 1963, **65**, 131–138.

(Received February 4, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 357–362

STIMULUS GENERALIZATION AS A FUNCTION OF DRIVE LEVEL, AND THE RELATION BETWEEN TWO MEASURES OF RESPONSE STRENGTH [1]

J. ROBERT NEWMAN AND G. ROBERT GRICE

System Development Corporation — *University of Illinois*

Generalization of an instrumental response in the rat to stimuli varying in the size dimension was investigated under 12 and 48 hr. of food deprivation. Generalization gradients for the 48-hr. condition were higher and of steeper slope than those for 12 hr. The data conform rather precisely to the implications of a multiplicative theory of habit and drive. The group means of 2 response measures, speed of the 1st test response, and number of extinction responses to the test stimuli, were linearly related with a correlation of .99. Implications of such findings for a reaction-potential concept are discussed.

It appears to be a rather general finding that the amount of stimulus generalization increases with increasing levels of drive. The situation concerning the slope of generalization functions as a function of motivational variables is somewhat less clear. Hull's multiplicative theory of habit and drive clearly predicts that these slopes should be steeper with higher drive levels under circumstances in which the response measure is linearly related to excitatory potential and ceiling effects are not involved. This is in direct opposition to the frequently expressed, commonsense view that high drive impairs discrimination and, hence, might be expected to flatten generalization gradients. The present experiment is an investigation of the effects upon size generalization, in the rat, of variations in food deprivation. The experiment was conducted in the situation previously used in generalization studies by Grice and Saltz (1950), Kling (1952), and Margolius (1955). Response speed and resistance to extinction are presented as two measures of response strength, and the relations between these two measures are reported.

[1] Data presented in this paper were collected in connection with a PhD thesis by the first author at the University of Illinois. The present report was supported by Grants G-14223 from the National Science Foundation and MH-08033-01 from the National Institutes of Health to the second author.

METHOD

Subjects.—The *Ss* were 120 experimentally naive albino rats of Sprague-Dawley strain. They ranged in age from 80 to 103 days at the start of the experiment. There were 65 females and 55 males divided as evenly as possible among the experimental groups. Nine additional *Ss* were eliminated from the experiment: 4 for failure during pretraining, and 5 for illness or death.

Apparatus.—The apparatus has been described in detail by Grice (1949). In brief, it consisted of a 2-ft. alley with a vertical sliding door in the middle. The stimulus objects were exposed at one end, and were mounted at the front of a 2-in. extension of the alley. The alley was pivoted at the center, so that each end became alternately the start box and the response compartment. The stimuli were white, circular, metal disks with closely fitting, square swinging doors in the center. The rat obtained a food pellet from a small tray attached to the back of the disk, by nosing open the door. The disks employed were 79, 50, 32, and 20 cm.^{2} in area, representing equal steps of log area.

Training procedure.—The training procedure has also been described in detail elsewhere (Grice & Saltz, 1950). On the first day, *Ss* were allowed to eat 20 pellets from the food tray with the swinging panel completely

open. On the second day, they received 10 trials, during which they were trained to push the panel by an approximation procedure. On each of the next 3 days, Ss received 20 training trials with the panel closed. Latencies were recorded from the raising of the start door at the center of the alley to the moving of the swinging panel by the animal. The 79-cm.[2] disk was used as the training stimulus for all animals.

During training, the hunger drive was produced by approximately 23½ hr. of food deprivation. Immediately following the training session each *S* received a ½-hr. feeding period, receiving Purina laboratory chow in an individual feeding cage. All *S*s were adapted to this schedule for 6 days prior to the start of the experiment. Food pellets used as rewards weighed approximately .05 gm.

Generalization tests.—Generalization tests were conducted 48 hr. following the third acquisition session, each animal being run at approximately the same time of day as in original training. Half of the *S*s were tested under 48 hr. of food deprivation and half under 12 hr. Each *S* received a 1-hr. feeding at the appropriate time before testing. This feeding included wet mash as well as the usual pellets.

The actual generalization testing was as described by Grice and Saltz (1950). Following 5 additional training trials to the original training stimulus 35 extinction trials were conducted with the test stimulus. If *S*s failed to respond in 30 sec., the trial was terminated and the next trial was run. Separate groups were run with each of the four stimuli under each drive level. There were 15 *S*s in each of the eight experimental groups.

Response measures.—Two measures of response strength to the test stimuli are presented. Resistance to extinction consists of the number of panel pushes during the 35 extinction trials. The other measure is speed of the first response—1,000/Latency. If *S*s failed to respond on the first test trial, time was cumulated over one or more additional trials until the first response. This manner of scoring latency is somewhat unusual, but it has previously been used successfully by Grice (1956). Following the reasoning of Estes (1950) and Spence (1954) in which latency is deduced as a function of momentary response probability, it should provide as good an index of reaction potential as the more conventional method. It also has the important advantage of permitting the obtaining of the extinction measure.

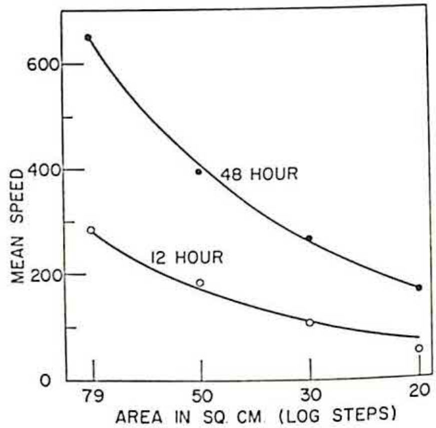


FIG. 1. Generalization gradients based on speed of the first-test response.

RESULTS

The mean speeds of the first test-trial response are presented in Fig. 1, and the mean numbers of responses in extinction to the test stimuli are presented in Fig. 2. The smooth curves drawn through the points and the transformation of Fig. 2 are discussed below. Analyses of variance of these data are presented in Table 1. The analyses generally support conclusions drawn from an inspection of the means. The greater levels of response strength with the 48-hr. deprivation are supported by the significant effects for Drive Level. The

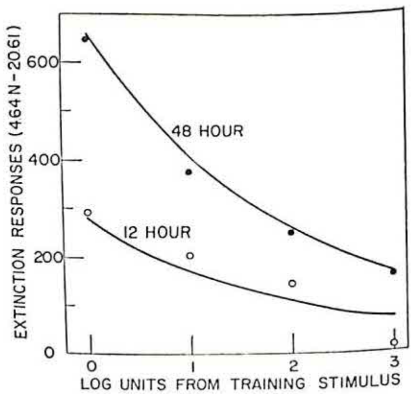


FIG. 2. Generalization gradients based on number of responses in 35 extinction trials to test stimuli.

TABLE 1

Analyses of Generalization Test Data

Source	df	F	
		Speed	Extinction
Drive (D)	1	26.62***	27.55***
Stimuli (S)	3	61.81***	19.83***
Stimuli within 48 hr.	(3)	7.91***	13.53***
Linear trend	(1)	22.58***	37.83***
Residual	(2)	0.57	1.38
Within stimuli	(56)	(84,807.2)[a]	(23.1)[a]
Stimuli within 12 hr.	(3)	8.69***	7.05***
Linear trend	(1)	25.68***	20.57***
Residual	(2)	0.19	0.29
Within stimuli	(56)	(17,518.3)[a]	(14.2)[a]
D × S	3	1.83	2.53
Difference in linear trend	(1)	5.06*	4.18*
Residual	(2)	0.26	1.44
Residual within drive	(116)	(50,288.0)[a]	(18.6)[a]
Within cells	112	(51,162.7)[a]	(18.6)[a]
Total	119		

[a] Error MSs.
* $p < .05$.
*** $p < .001$.

significant Stimulus effects and the significant linear trends indicate that generalization decrement occurred under both levels of drive. The significant effects for Difference in Linear Trend suggest that the apparently steeper slope for the high-drive condition is genuine. The conclusions are essentially the same for both response measures.

Discussion

Generalization and drive strength.—The results of this experiment agree with the implications of the multiplicative theory of habit and drive that generalization gradients should be both higher and steeper under high-drive conditions than at lower levels of drive. In view of the fact that response speed is generally regarded to be linearly related to reaction potential (E), a further analysis of these data has been made to investigate the degree to which the multiplicative theory is supported. Assuming an exponential form of the gradient, the following expression may be written:

$$E = D(H \times 10^{-bd})$$

E indicates excitatory potential; D, drive; H, habit strength to the training stimulus; d, the number of logarithmic units between the training and test stimuli; and b is an empirical constant. Letting D equal unity for the 48-hr. condition, the following equation was fitted to these data:

$$E_{48} = 658 \times 10^{-0.198d}$$

This equation is plotted as the upper curve in Fig. 1. Under the multiplicative theory, the 12-hr. data should be approximated by multiplying this expression by a lower value for D, less than unity. Such a parameter was estimated from the 12-hr. data, and the lower curve in Fig. 1 represents the above equation multiplied by the lower value, 0.43. The new equation becomes:

$$E_{12} = 283 \times 10^{-0.198d}$$

The fit to the 12-hr. data is good and supports the implication that the only difference between the two functions is difference in a multiplicative constant dependent upon drive level.

Including the present experiment there now exists considerable support fo

the implication of the multiplicative theory that generalization gradient slope and drive level should be positively related. In addition to this experiment, the prediction is supported by the Porter (1962) experiment, and by the unpublished[2] absolute generalization data of Jenkins, Pascal, and Walker (1958). As was pointed out by Porter, data for three of the four drive levels employed by Thomas and King (1959) also conform to the prediction. The discrepant group in that study was that with the highest drive level used, a group of pigeons reduced to 60% of initial body weight. One interesting fact about the data for this group, is that the origin of the gradient at the training stimulus is actually below that for the next lowest drive level, a 70% group. This suggests the possible interpretation that this group may have been suffering from an inanition factor which depressed the maximum response rate. This would account for the flatter function obtained for this condition on the basis of a ceiling effect, imposed by the inanition factor, depressing the central region of the gradient. It is also possible that the difference in the slopes of the *relative* generalization functions obtained by Jenkins, Pascal, and Walker could be explained in the same manner. Adequate evaluation of this theory is not an uncomplicated matter. For example, it should be remembered that Hull (1951) did not assume that response potentiation was a monotonic function of deprivation conditions. Before attempting to investigate the quantitative effects upon generalization functions, one should be aware of the relations between the response measures used and the motivational variable for ungeneralized responses. It must also be recognized that the role of such factors as inanition may be expected to interact with such variables as species, strain, age, maintenance conditions, and condition of health of *S*s.

There appears to be little support of the view, based on the assumption of impaired discrimination under high drive, of an inverse relation between generalization gradient slope and drive level. However, it seems clear that such results can be obtained under conditions in which the maximum levels of the response measures are approached. Such maxima may either be inherent in the measures, or they may be introduced by experimental conditions which impair *S*'s ability to respond. Perhaps flattening of gradients may also be obtained in cases of sensitization or pseudoconditioning, where response evocation may be nonassociatively determined.

Relations between the response measures.—One result of some interest in the present experiment is the close agreement between the two response measures used in the generalization tests. There have been occasional references to rather low correlations between various measures of response strength; e.g., Kimble (1961). Sometimes such results are regarded as an argument against the utility of an excitatory or reaction-potential concept. Following a procedure previously reported by Grice (1956), the correlation between the present two measures has been analyzed into within-conditions and between-conditions components. The obtained correlations between speed of the first response and responses in extinction were .22 for within conditions, but .99 for between conditions. This latter value is, of course, the correlation between the group means. The between-conditions correlation was also .99 between speed of the first response and the percentage of responses on trials remaining after the first response. The smooth curves drawn through the extinction points of Fig. 2 are not fitted to the data, but are actually the same curves from the analysis of the speed data. The linear transformation of the ordinate is the linear regression equation.

Correlations between response measures usually reported are based on a group of *S*s, similarly treated, and are analogous to the present within-conditions correlation. A possible interpretation is that when a group of *S*s is subjected to extended training, systematic individual differences in reaction

[2] W. O. Jenkins, personal communication, 1962.

potential are small in relation to behavioral oscillation and measurement error. However, when systematic differences in reaction potential are introduced between groups by the operation of experimental variables, high correlations may be obtained. The between-conditions correlation of .99 leaves little room for doubt that the measures are measuring the same thing, and that they are linearly related within the range included. A plot of the relation between first-trial speed and percent response on remaining trials is presented in Fig. 3. The main import of the figure is not merely the close relationship, which is clearly indicated by the correlation coefficient, but is the illustration of the positions of the various group means. It is of considerable interest to note that the relative position of a given point is determined jointly by the amount of generalization decrement and by the amount of deprivation. The fact that all points fall along the same regression line strongly suggests that both experimental variables are influencing the same thing.

Clearly, it is not to be expected that such a simple picture will always be obtained. For example, there is evidence that, in the runway, amount of training and resistance to extinction are not positively related (Ison, 1962). On the other hand, the present data do not stand entirely alone. In an experiment with constant drive, but with generalization in both directions, Grice (1956) has reported a between-conditions correlation of .89 over a somewhat more attenuated range of the same two variables. Perin (1942) reported extinction data and initial latency in an experiment involving concurrent manipulation of both deprivation and amount of training. A plot of extinction trials and median initial speed (reciprocal latency), indicates that these data also fall on a single regression line, and the between-conditions correlation is .94. Taken together, these three sets of data suggest that, at least within a limited domain, deprivation, generalization decrement, and amount of training influence a common underlying variable, which may be measured by either response speed or resistance to extinction. Specifying the boundaries of this domain would appear to be a problem of considerable theoretical importance.

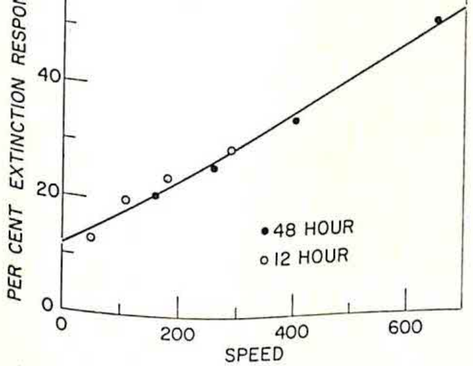


FIG. 3. Between-conditions relation between speed of the first-test response and percent responses on remaining test trials. (The straight line is from the linear regression equation.)

REFERENCES

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

GRICE, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, **42**, 365–373.

GRICE, G. R. Response speed and size generalization. *Psychol. Rep.*, 1956, **2**, 246.

GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

ISON, J. R. Experimental extinction as a function of number of reinforcements. *J. exp. Psychol.*, 1962, **64**, 314–317.

JENKINS, W. O., PASCAL, G. R., & WALKER, R. W., JR. Deprivation and generalization. *J. exp. Psychol.*, 1958, **56**, 274–277.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

KLING, J. W. Generalization of extinction of an instrumental response varying in the size dimension. *J. exp. Psychol.*, 1952, **44**, 339–346.

MARGOLIUS, G. Stimulus generalization of an instrumental response as a function of

the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.

Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

Porter, J. J. Stimulus generalization as a function of UCS intensity in eyelid conditioning. *J. exp. Psychol.*, 1962, **64**, 311–313.

Spence, K. W. The relation of response latency and speed to the intervening variables and N in S-R theory. *Psychol. Rev.*, 1954, **61**, 209–216.

Thomas, D. R., & King, R. A. Stimulus generalization as a function of level of motivation. *J. exp. Psychol.*, 1959, **57**, 323–328.

(Received February 5, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 363–368

STIMULUS GENERALIZATION AS A FUNCTION OF DRIVE SHIFT[1]

ROBERT B. ZAJONC AND DAVID V. CROSS

University of Michigan

6 pigeons reduced to 70% and 6 reduced to 90% of their body weight were trained to discriminate a 10-mm. circular spot from 4- and 16-mm. spots. Following training, Ss in the 1st group were raised to 90% and Ss in the 2nd dropped to 70% of their body weight. All birds were tested for generalization to spots 6, 8, 10, 12, and 14 mm. in diameter. Following brief discrimination retraining the birds were returned to their original body weights and again tested for generalization. The generalization gradients were found to shift toward larger stimuli when drive level was raised and toward smaller stimuli when it was lowered.

A series of recent studies have shown that a change in drive, following training, produces a lateral displacement of the generalization gradient (Zajonc & Dorfman, 1964). Platz (1962) found such a displacement in generalization along the loudness continuum as a function of changes in conditioned fear. Similar displacements were found when drive was replaced by irrelevant extraneous stimulation. For instance, Dorfman (1961) found a lateral displacement in generalization to tactual stimuli as a function of changes in the amplitude of auditory stimulation occurring simultaneously with the eliciting stimuli, and Karp (1961) corroborated Dorfman's findings using the same stimulus continuum and the same extraneous stimulation but a somewhat different training procedure. These results together with some findings in the area of sensory interaction were taken by Zajonc and Dorfman (1964) to indicate that drive not only functions as an energizer of responses, but that it has perceptual effects as well. By assuming that the stimulus component of drive interacts with the perception of the eliciting stimulus, either increasing or decreasing its apparent intensity, these authors were able to account for the lateral displacements of generalization gradients following a drive shift.

The experiments referred to above used human *S*s to examine the effects of drive shift on stimulus generalization. In the present study pigeons were used to investigate size generalization as a function of changes in the schedule of food deprivation.

METHOD

Twelve male, White Carneaux pigeons were used as *S*s. All *S*s were placed on restricted feeding until their body weights reached 80% of their free-feeding level.

Apparatus.—A conventional experimental chamber for pigeons (Grayson-Stadler, E1100PB) was used. It was modified by mounting a small stripfilm projector on its side so that the stimulus (the only source of light in the chamber) could be focused on the pecking key. Seven stimuli were employed during the course of the experiment. These consisted of white circular spots on black backgrounds photographed and copied on 35-mm. high contrast copy film. The stimuli, as projected onto the back of the pecking key, ranged in size (in 2-mm. steps) from 4 to 16 mm. in diameter.

Procedure.—Following the achievement of stable weight, all *S*s were magazine trained and conditioned to keypeck with the key con-

[1] This study was part of a research program supported by the Air Force Office of Scientific Research, Grant AF 49(638) 367. We wish to thank David Carter for his assistance in the tabulation and analysis of results.

tinuously illuminated by a 10-mm. diameter spot of light. After one session during which 50 continuous reinforcements were provided, the pigeons were divided equally into two groups: Group I *S*s were raised to 90% of their free-feeding body weight and Group II *S*s were dropped to 70% of their free-feeding level. Controlled feeding, during all of the training sessions that followed, maintained the body weight of each *S* within 5 gm. of that established. All *S*s received 60 min. of fixed interval 15-sec. training to the 10-mm. spot, consisting of 100 15-sec. work periods separated by 15-sec. blackouts. Reinforcement consisted of access to the food magazine for 5 sec. All *S*s were then given nine daily sessions of discrimination training during which each of three stimuli, (the initial 10-mm. spot and the 4- and 16-mm. spots) were presented 16 times in random sequence for 15-sec. periods separated by 15-sec. blackouts. Responding to the 10-mm. spot was reinforced at the end of each presentation period with 15 sec. access to the food magazine, and responding to the 4- and 16-mm. spots was extinguished. The *S*s were run in pairs, one *S* from each group, matched for initial response rate. Each pair was run at the same time every alternate day. Following the last day of discrimination training each *S* was shifted in body weight; Group I *S*s were shifted to 70% of their free-feeding level and Group II *S*s were shifted to 90% of their free-feeding level. Three to 4 wk. were permitted to elapse between training and testing for each pair until the new levels were stabilized. Generalization testing consisted of repeated presentation of five stimuli (6-, 8-, 10-, 12-, and 14-mm. spots) under extinction conditions. The 4- and 16-mm. spots were omitted. Each stimulus was presented 10 times for 15 sec. according to a predetermined schedule of random permutations of the five stimuli. Fifteen-second blackouts separated each presentation. Each *S* of a given pair received the same random sequence of trials but different random sequences were used for different pairs. Three daily test sessions were given. At the end of each session the 10-mm. spot was presented briefly (5 sec.) and responding was reinforced. Following these testing sessions, body weights were again shifted to the level previously established for discrimination training and one additional session of generalization testing was given. Thus, Group I received training under low-drive conditions and generalization testing under high-drive conditions, followed by further testing under low drive. Group II received identical treatment under high drive, low drive, and again high drive.

Results

Figure 1 shows the effects of stimulus size and drive level on response frequency during each of 3 consecutive days of generalization testing. The Group I curves represent total responding to each stimulus value for all of six birds trained under low-drive and tested under high-drive conditions. Although all birds in Group II (trained under high and tested under low drive) were given 3 days of testing, some failed to perform after the first or second day. Thus, the Group II curves represent response totals for six, three, and two *S*s for Days 1, 2, and 3, respectively. It appears that the shape of the generalization curves for each group remained stable over test sessions. It can also be seen that birds for whom drive was increased during testing responded to stimuli 12 and 14 mm. in diameter at a higher rate than to stimuli 6 and 8 mm. in diameter. For birds trained under high and tested under low drive, the converse is true. An analysis of variance of these response measures showed that both main effects, stimulus size, $F\ (4, 50) = 30.342, p < .001$, and drive level, $F\ (1, 50) = 58.048$, $p < .001$, were significant as was the effect due to their interaction, $F\ (4, 50) = 4.511$, $p < .01$. These results are reported primarily to establish the fact that drive had a significant effect on overall rate of responding. Our chief interest, however, is in the lateral displacement of the generalization functions. To control for the effects of drive on overall response rate, the above measures were normalized relative to the total responding for each *S*. The generalization functions relating proportion of total responses to stimulus size for the two groups are shown in Fig. 2. Analysis based on these proportions provides a

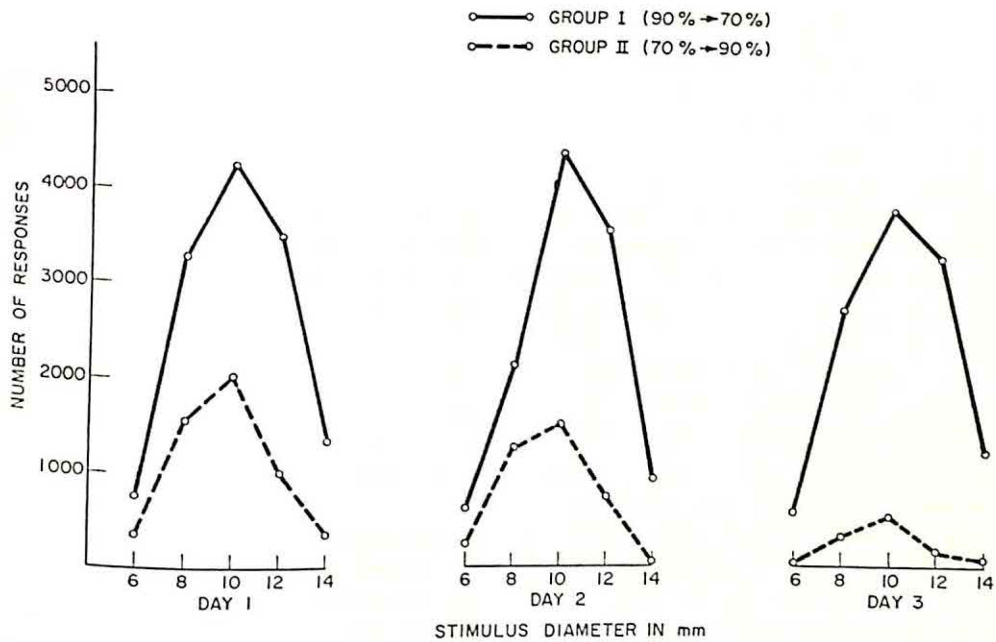


FIG. 1. Stimulus generalization gradients in terms of absolute response rate for 3 consecutive days following the first drive-level shift.

better test of our hypothesis which was concerned with drive-produced directional asymmetries in the generalization functions. An orthogonal polynomial trend analysis was performed. The results are summarized in Table 1. A significant interaction effect was obtained ($F = 6.302$, $p < .001$) owing to differences in the asymmetric (linear and cubic) components of the two generalization functions. These differences are exhibited in Fig. 2 in the relative displacement to the *right* in the Group I gradients and to the *left* in the Group II gradients.

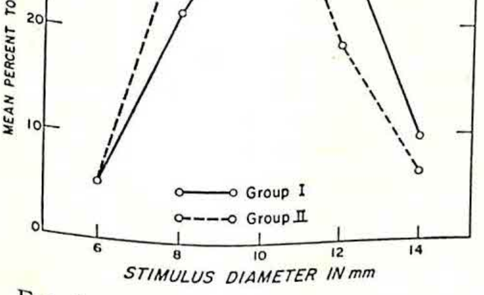


FIG. 2. Stimulus generalization gradients in terms of percentage of total responses following the first drive-level shift.

The results for individual *S*s are shown in Fig. 3. The generalization curve obtained from each *S* is compared with that of the *S* with which it was paired in the experimental treatment. The response measures were normalized relative to the total number of responses emitted by each *S*. The interaction effect is exhibited in the generalization behavior of all pairs of *S*s. There is a modal shift for one bird trained under low and tested under high drive and in the opposite direction for two birds which were trained under high and tested under low drive. The gradients of birds

tested under low drive are somewhat steeper than those for birds tested under high drive, a finding consistent with the results obtained by Thomas and King (1959).

When *S*s were shifted back to the body weight at which they were initially trained and then retested for generalization, a tendency to reverse the initial pattern of behavior was observed. Three *S*s failed to perform to a criterion of the total emission of at least 500 responses during this session, and were thus not included in the calculation of results. Since Group I was originally tested under 70% of body weight, the retest for this group involved a *decrease* in drive level. Group II, on the other hand, was originally tested under 90%, and the retest for these animals represented an *increase* in drive. The displacement of the gradients as a result of first and second drive shift is best seen when we compare normalized response rates to stimuli 6 and 8 mm. in diameter with those to stimuli 12 and 14 mm. in diameter. The percentage of total responses (excluding those emitted to the middle stimulus, 10 mm.) is shown in Table 2. Whenever drive is increased the response rate to the two smallest stimuli is less than that to the two largest stimuli. This is true for the first test in Group I and retest in Group II. A decrease in drive, on the other hand, leads to an increased response rate to the two smallest and decreased response rate to the two largest stimuli (Group II). In Group I, however, while the response rate to the smallest stimuli is higher and to the largest is lower on retest than it was on the first test, the second shift is quite weak.

Considering that 2 mo. elapsed

TABLE 1

ANALYSIS OF VARIANCE FOR DATA IN FIG. 2

Source	*df*	*MS*	Error Term (Row)	*F*
Stimulus effect	(4)	(201,615.75)	E	82.29***
1. Linear	1	1,197.01	E.1	.46
2. Quadratic	1	781,053.65	E.2	41.77***
3. Cubic	1	6,931.20	E.3	2.55
4. Quartic	1	17,281.07	E.4	6.58*
Between group means (Drive Effect)	(1)	0		
Between group trends (Drive Effect × Stimulus Effect)	(4)	(14,779.00)	E	6.03***
1. Linear	1	18,775.00	E.1	7.27*
2. Quadratic	1	3,146.05	E.2	1.68
3. Cubic	1	34,408.50	E.3	12.65**
4. Quartic	1	2,786.70	E.4	1.06
Between individual means (Response Rate)	(10)	0		
Between individual trends	(40)	(2,449.97)		
1. Linear	10	2,583.47		
2. Quadratic	10	1,869.76		
3. Cubic	10	2,720.21		
4. Quartic	10	2,626.43		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

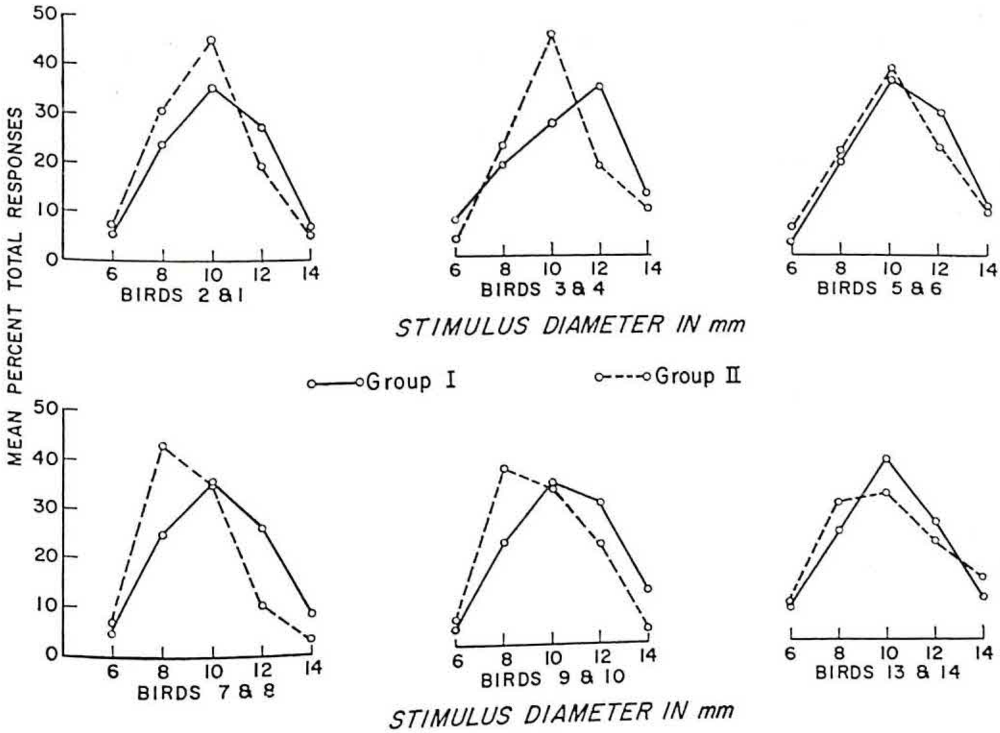


FIG. 3. Stimulus generalization gradients in terms of percentage of total responses following the first drive-level shift for individual pairs of matched *S*s.

since the last day of training and that these birds were tested under low-drive conditions subsequent to 3 days of extinction under high drive, it is understandable that their behavior would exhibit noticeable variability. In fact, the retest generalization functions in Group I were displaced to the left for two *S*s, and to the right for the other two.

TABLE 2

MEAN PERCENTAGE OF RESPONSES TO THE LEFT AND RIGHT WINGS OF GENERALIZATION GRADIENTS

Wings	Group I		Group II	
	Test 1 70%	Retest 90%	Test 1 90%	Retest 70%
Left (Stimuli 6 and 8 mm.)	41.5	44.1	58.5	38.8
Right (Stimuli 12 and 14 mm.)	58.5	55.9	41.5	61.2

DISCUSSION

In the experiments quoted above by Dorfman (1961), Karp (1961), and Platz (1962) an increase in drive resulted in a shift of the generalization gradient toward stimuli greater in intensity than the conditioned stimulus, while a decrease in drive resulted in a lateral displacement toward smaller stimulus intensities. It is clear from the results of the present experiment that a lateral displacement of the generalization was also obtained as a function change in the level of hunger. However, in contrast with the previous findings, an increase in drive shifted the gradients in the direction of larger circles and a decrease in drive in the direction of smaller circles. A lateral shift in generalization gradient of a similar character was also found by Matlin (1960). In Matlin's experiment *S*s were

trained to make one response to a tone of medium intensity (77 db.) and another to one tone of lower (70 db.) and one tone of higher intensity (85 db.). Dynamometer pressure served to manipulate drive level. During training one group of *S*s pressed a dynamometer to $\frac{2}{3}$ of their maximum and another to $\frac{1}{20}$ of their maximum. During testing observations were made of generalized responses to the three training and four other stimuli (73, 75, 79, and 81 db.). The generalization trials involved lowering drive for the former and increasing drive for the latter group of *S*s. A significant lateral displacement of generalization gradients toward stimuli of greater intensity was obtained for the group trained under low and tested under high drive. A tendency in the opposite direction was observed for the group trained under high and tested under low drive. On the other hand, Carter and Zajonc (1962) in a similar attempt trained two groups of pigeons to keypeck for food reinforcement. One group was given such training during the presence of a constant electrical stimulation high in intensity (.8 ma.) and another low in intensity (.4 ma.). Responses were reinforced only in the presence of a 1,000-cps, 43.0-db. tone. Following training, shock intensity was increased for the low-shock group and decreased for the high-shock group and all *S*s were tested for generalization to seven auditory stimuli of 1,000 cps varying in sound pressure from 27.0 to 69.5 db. As in the other studies no reinforcement was administered during generalization testing. These investigators found that the generalization along the loudness dimension is displaced toward stimuli less loud than the CS when drive is increased following training, and toward stimuli louder than the CS when drive level is decreased. These shifts were obtained, however, only on the initial trials following drive shift.

In the light of these findings it is apparent that drive does in fact interact with stimulus generalization, and the possibility of perceptual effects deriving from the stimulus component of drive is strengthened by the present results. However, no simple rule seems to describe the nature of this interaction, nor to predict the *direction* of the lateral shift.

REFERENCES

Carter, D. E., & Zajonc, R. B. Studies on drive and incentive in perception VI: Some effects of drive produced by electric stimulation on generalized responses to loudness. Technical Report No. 16, 1962, Research Center for Group Dynamics, Ann Arbor, Michigan, Contract AF 49(638)367.

Dorfman, D. D. Some effects of drive on the effective stimulus intensity. *Psychol. Rev.*, 1961, 9, 87–98.

Karp, S. E. Studies on drive and incentive in perception: III. Some effects of drive on stimulus generalization obtained with a revised training procedure. Technical Report No. 8, 1961, Research Center for Group Dynamics, Ann Arbor, Michigan, Contract AF 49(638)367.

Matlin, A. H. Studies on drive and incentive in perception: II. The effect of drive produced by proprioceptive stimulation on generalized responses to loudness and pitch. Technical Report No. 6, 1960, Research Center for Group Dynamics, Ann Arbor, Michigan, Contract AF 49(638)367.

Platz, A. *Some effects of conditioned fear on stimulus generalization.* (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1962, No. 63–6938.

Thomas, D. R., & King, R. A. Stimulus generalization as a function of level of motivation. *J. exp. Psychol.*, 1959, **57**, 323–328.

Zajonc, R. B., & Dorfman, D. D. Perception, drive, and behavior theory. *Psychol. Rev.*, 1964, **71**, 273–290.

(Received February 10, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 369–374

SUBJECTIVE INTENSITY OF MINERAL TASTE IN WATER[1]

WILLIAM H. BRUVOLD AND WILLIAM R. GAFFEY

University of California, Berkeley — *California State Department of Public Health*

An experiment determined the ability of a power function to summarize ratings given solutions of NaCl, $MgCl_2$, and Na_2SO_4. 16 Ss rated 9 solutions of a salt using a magnitude-estimation procedure. A separate experiment was devoted to each of the 3 salts studied. Results indicated that a power function, $\psi = k^n$, could account for a substantial portion of the variance in the ratings given. Further analyses of individual ratings showed that power functions varied significantly between Ss. It was proposed that the individual variation could be accommodated by an analysis of variance model.

The method of magnitude estimation, described by Stevens (1956, 1958), has been used to obtain ratings of the subjective intensity of stimuli falling along a prothetic continuum. Stimuli that form a prothetic continuum differ in physical intensity, but they all belong to a single qualitative dimension. A substantial number of different prothetic continua have been investigated using magnitude-estimation procedures. A general finding has been that the relationship between the physical intensities of the stimuli and their reported subjective intensities can be described by the equation $\psi = k\varphi^n$ (Stevens, 1960) or by the equation $\psi = k(\varphi - \varphi_0)^n$ (Stevens, 1962).

Only a small amount of research has been performed that employs magnitude estimation to investigate subjective intensity for the modality of taste. Stevens (1960) reports, without detailing the exact methods employed, that McLean found exponents (n in the above equations) of 0.8 for saccharine, 1.3 for sucrose, and 1.3 for sodium chloride. Jones and Marcus (1961) determined exponents for 98 ratings of sodium chloride solutions. The median exponent of the 98 computed was 1.05. These two researches appear to report the major data for tasted substances generated by magnitude estimation and described by a power function.

The present study was conducted in order to obtain information concerning *S*s' ratings of the subjective intensity of the taste of solutions containing different amounts of a mineral solute. Three separate experiments were performed using the method of magnitude estimation. Sodium chloride was used as the solute in the first experiment. Magnesium chloride and sodium sulphate were used as solutes in the second and third experiments, respectively. The specific procedures employed in the experiments were designed to provide a test of the adequacy of the power law equation, to analyze individual differences in ratings, and to compare the power functions representing the three minerals rated.

METHOD

Solutions

Minerals commonly found dissolved in water include calcium, sodium, and magnesium in combination with chloride, sulphate, or carbonate (Babbitt, Doland, & Cleasby, 1962). After considerable pilot work

[1] The research reported in this paper was supported in part by Fellowship Award WSP–16, 452(1) from the Public Health Service.

was completed, it was decided to use NaCl, $MgCl_2$, and Na_2SO_4 as solutes for study. None of these compounds has any odor, and all have a high level of solubility.

In each experiment a series of nine solutions was employed. Using distilled water as the solvent and a reagent grade chemical as solute, solutions of 1,000; 2,000; 4,000; 7,000; 10,000; 20,000; 40,000; 70,000; and 100,000 mg. per liter (ppm) of NaCl were prepared for use in the first experiment. Series of solutions containing the same solute levels were prepared, when needed, for use in the $MgCl_2$ and Na_2SO_4 experiments. Both NaCl and Na_2SO_4 were of anhydrous form, and their exact weights were used in preparing the nine solutions. The $MgCl_2$ was hydrated, having the formula of $MgCl_2 \cdot 6H_2O$. Proper allowance was made for the hydration factor in preparing the $MgCl_2$ solutions. The error, in ppm of dissolved solute, for each solution in each series was less than 1% at the time the solutions were delivered from the chemical laboratories.

Subjects

Sixteen *S*s served in all three experiments. Of these, 12 were males and 4 were females. The ages for the group ranged from 16 to 51 yr. All *S*s served voluntarily without pay.

The sample of *S*s employed was heterogeneous in many respects. Age and sex variations have been noted. Educational level varied from students in high school to recipients of doctoral degrees. Occupation also varied widely. Stenographers, chemists, high-school students, professors, engineers, and housewives were included in the sample. Perhaps the best general description that could be given of the entire group of *S*s would be to state that all resided in the San Francisco Bay Area and all had some connection with the University of California or the California State Department of Public Health.

Procedure

Before *S* arrived for an experimental session, *E* prepared two series of solutions. Each series contained the nine solute levels described above. Six standard solutions of 10,000 ppm were also prepared. All solutions were presented in 50-ml. beakers, filled to the 30-ml. level. The solutions to be rated bore no identification save a single letter to allow *E* to keep proper records. The standard solutions were all marked with a capital "S," and were clearly and easily identifiable.

Each experimental session was divided into two parts, with a 5-min. rest between parts. During each part, *S* rated nine solutions against the standard solution. A standard was presented first, followed by three unknown solutions. This block of solutions was presented three times. Thus, Part 1 of an experimental session consisted of three blocks of four solutions each, where each block contained a standard and three unknown solutions. After a 5-min. rest interval, the same procedure was repeated. The order of the standard solutions was known to *S* and fixed. The order of the solutions to be rated was unknown to *S* and determined by chance.

The experimental sessions were conducted in a small air-conditioned room, the temperature of which was constantly kept very near to 70° F. A laboratory table with a sink was used for the sample presentation and expectoration. A large electric clock was visible to *S* for use in timing rest periods.

Instructions given to *S* before the start of each experimental session served to explain the rather complicated duties he was to perform. The entire experimental procedure, including the full set of instructions, was repeated three times at intervals of about 30 days for each *S*.

Instructions

The instructions given to each *S* at the beginning of each experimental session are presented below.

[General Instructions] This is an experiment to scale the subjective intensity of salt solutions. You will be asked to give numbers that reflect how strong the solution tastes to you in accordance with the instructions. The standard solution is called 100. That is, it is given the number 100 to provide an "anchor" for a scale. Your job is to rate the remaining 9 salt solutions in terms of the standard and its number. If the 1st solution tastes twice as strong to you as the standard, call it 200. If it tastes half as strong, call it 50. If it tastes 16 times as strong, as the standard call it 1600. If it tastes $\frac{1}{4}$ as strong, call it 25. TRY TO ASSIGN THE NUMBERS SO THAT THEY INDICATE HOW STRONG THE SOLUTION TASTES IN RELATION TO THE STANDARD. Do not feel overly constrained by numbers. Let your taste be the major factor in determining the number that reflects how the various solutions compare with the standard.

[Phase I] The exact procedure you are

to follow during Phase I of the experiment is outlined below.

1. Take *all* of the standard into your mouth and hold it for about 5 seconds and then spit it into the sink.

2. Fill the larger beaker with tap water. Rinse your mouth thoroughly with the tap water. Repeat the rinsing until the salty after-taste disappears. Do not rinse more than 5 times. Wait approximately 1 minute.

3. Take *all* of the first solution into your mouth and hold it for about 5 seconds and then spit it into the sink.

4. Report your rating of the intensity of the solution to the experimenter.

5. Rinse your mouth thoroughly as indicated above. Wait approximately 1 minute.

6. Repeat steps 3, 4, and 5 until all 9 solutions have been rated. Remember to rinse and wait 1 minute between tasting each solution.

[Phase II] During Phase II of the experiment you are to rest for 5 minutes. Sit quietly in the chair during this time. Do not leave the experimental room.

[Phase III] During Phase III you are to repeat exactly the procedures followed during Phase I. These procedures can be summarized as: 1) Taste the standard and rinse; wait 1 minute. 2) Taste the 1st solution and report the numerical rating. 3) Rinse and then wait 1 minute. 4) Repeat these steps until all 9 solutions have been rated.

In addition to these instructions, each *S* was given a printed summary of the six steps outlined under Phase I above. The summary was placed on the laboratory table in front of *S* to remind him of the details of the rating procedure. Multiples and fractions of 100 were included on the summary sheet. It was indicated in bold print that these multiples and fractions were to serve as examples only, and that they were not to be construed as numbers that had to be used in the ratings.

RESULTS

The logarithm of a response which follows the power law can be expressed as a linear function of the logarithm of the stimulus plus a random error. If the power law does not hold, the relationship between the logarithms can still be expressed as the power law, plus deviations from the power law, plus a random error. Thus, the log of the response of any one individual for a particular stimulus can be written as

$$\log \psi_{ij} = \log k + n \log \varphi_i + \delta_i + E_{ij}$$

where ψ_{ij} is the response magnitude at the jth replication for the stimulus of the ith intensity, δ_i is the deviation from the power law at the ith stimulus intensity, and E_{ij} is random error. In the present study there were nine levels of stimulus intensity ($i = 1 \cdots 9$), measured by the amount of solute in solution, and 2 ($j = 1, 2$) replications at each intensity level. The power law applies if and only if all $\delta_i = 0$, that is, if the regression of the log response on the log stimulus intensity is linear. It was assumed that E_{ij} was normally distributed, homoscedastic, and mutually independent. The third condition is assured by the experimental design, and failure of the first two does not substantially affect analyses of variance (Scheffé, 1959).

Least-squares estimates of log k and n, residual sums of squares, and sums of squares around regression were calculated for each of the 16 *S*s for NaCl solutions. The same calculations were also made for $MgCl_2$ and Na_2SO_4 solutions. Thus, a total of 48 separate analyses of regression was

TABLE 1

POWER LAW CONSTANTS

Constant	Solute	Group Value	Range[a]
k	NaCl	.021	.000– .289
	$MgCl_2$	.021	.000–2.228
	Na_2SO_4	.027	.000–1.926
	All solutes	.023	.000–2.228
n	NaCl	.920	.634–1.564
	$MgCl_2$	.937	.498–1.598
	Na_2SO_4	.892	.481–1.421
	All solutes	.916	.481–1.598

[a] ".000" means zero to the nearest thousandth.

curve are simply estimates of log k (and thus of k) and n, the basic population constants.

Jones and Marcus (1961) reported that an individual-modality interaction for exponents (analogous to n_{is} in the above equation) could be accounted for by a product involving individual and modality main effects. The occurrence of a significant individual-solute interaction for exponents obtained in the present study cannot be similarly explained as a product involving individual and solute main effects since the solute main effects were found to be nonsignificant. The application of Tukey's test for nonadditivity confirms this reasoning by finding no significant multiplicative component in the individual-solute interaction. Apparently the interaction arises from idiosyncratic reactions of *S*s to particular solutes. The reactions increase the response for some *S*s and decrease it for others, so that the overall average remains essentially constant from one solute to the next.

It would be instructive to investigate the extent to which interaction of the type found by this study is due to casual divergences in the individual environments of *S*s, as opposed to more fundamental differences. In the former case, a repetition of the experiment with a homogeneous population, and with a shorter time between tests of the different solutions, might be expected to reduce the interaction sharply.

REFERENCES

Babbitt, H. E., Doland, J. J., & Cleasby, J. L. *Water supply engineering.* (6th ed.) New York: McGraw-Hill, 1962.

Jones, F. N., & Marcus, M. J. The subject effect in judgments of subjective magnitude. *J. exp. Psychol.*, 1961, **61**, 40–44.

Scheffé, H. *The analysis of variance.* New York: Wiley, 1959.

Stevens, S. S. The direct estimation of sensory magnitudes: Loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196.

Stevens, S. S. The psychophysics of sensory function. *Amer. Scientist*, 1960, **48**, 226–253.

Stevens, S. S. The surprising simplicity of sensory metrics. *Amer. Psychologist*, 1962, **17**, 29–39.

(Received February 11, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 375–381

OVERLEARNING AND BRIGHTNESS-DISCRIMINATION REVERSAL [1]

M. R. D'AMATO AND DONALD SCHIFF

Rutgers University *State University College at New Paltz*

In 8 experiments involving a total of 230 rats the overlearning-reversal effect failed to occur in a brightness-discrimination task despite the manipulation of a number of possibly relevant variables. The usual pattern of results was a small, nonsignificant, difference in numbers of trials to reversal criterion, favoring the controls as often as the overtrained Ss.

The literature on the overlearning-reversal effect—the facilitation of reversal learning by extensive postcriterial training—is perplexing. After Reid's (1953) initial demonstration that rats allowed extensive overtraining on a brightness discrimination would reverse faster than *S*s reversed immediately after reaching criterion, there subsequently appeared a number of confirming reports. For a time the overlearning-reversal effect (ORE) seemed so well-established that it found employment as an analytic tool in comparative psychology (e.g., Brookshire, Warren, & Ball, 1961; Warren, 1960; Warren, Brookshire, Ball, & Reynolds, 1960). At a somewhat more general level, the facilitative effects of overtraining, of which the ORE is but a special case, has had important theoretical consequences (e.g., Mandler, 1962).

One of the first indications that the ORE might not be the general phenomenon it appeared to be came in a report of four separate failures to find an ORE in a position-reversal task (D'Amato & Jagoda, 1962). In very short order these failures were verified by others (Clayton, 1963; Hill & Spear, 1963; Hill, Spear, & Clayton, 1962; Komaki, 1962), bringing to at least 11 the number of experiments in which the ORE could not be obtained in a position-discrimination task. On the positive side there is the initial success by Pubols (1956) and a more recent report by Capaldi (1963), though there may be others that have escaped us (cf. also Brookshire et al., 1961). If the ORE is a genuine phenomenon in position-discrimination reversal, the set of conditions under which it can be expected to occur must be very restricted indeed.

Still, the major evidence behind the ORE comes from visual-discrimination tasks, usually black-white discriminations, and as far as we know there is only one published report in which the ORE failed to appear in the visual-discrimination setting (Erlebacher, 1963). But here too, after an initial marginal success (D'Amato & Jagoda, 1961), our efforts to lay hold of the ORE within a brightness-discrimination task have repeatedly met with failure. Because the phenomenon seemed so well-established we were reluctant to publish our results until we had unearthed the variables responsible for our unsuccessful attempts. Having now exhausted all promising hypotheses as to the causes of our fruitless search, a record of our

[1] This research was supported by Research Grants G-14724 and GB-1515 from the National Science Foundation. H. Jagoda and D. Keller assisted in various phases of the experiments.

studies may be of value to those who wish to concern themselves with this important, interesting, but apparently elusive, phenomenon.

Experiment I

It should be pointed out that the ORE, when it does occur, is an effect of considerable magnitude, statistically demonstrable with small numbers of *S*s. Consequently, we adopted the strategy of using relatively small groups of *S*s in our sequence of experiments, intending to follow up promising results with larger scale studies.

The first experiment was essentially a repetition of our initial brightness-discrimination study (D'Amato & Jagoda, 1961), aimed at determining whether the relatively small difference obtained in that study (between Group C and Group C-200) was simply a chance result.

Method

Subjects.—The *S*s in Exp. I–VI, naive albino rats (45% males), were taken from litters bred in our laboratory under standardized conditions. They ranged from 75 to 140 days of age when beginning pretraining.

Apparatus.—Four single-unit automatic Y mazes, described in detail in a previous report (D'Amato & Jagoda, 1960), were used in this and subsequent experiments.

Pretraining.—Four days prior to the beginning of discrimination training *S*s were placed on a 23-hr. water-deprivation schedule, food being available ad lib. The only pretraining given in the Y mazes consisted in permitting *S* to drink five dippersful of water on the day prior to the beginning of training; three dippersful were allowed on the subsequent day. This pretraining procedure was followed in all experiments.

Discrimination training.—All *S*s were run 20 trials per day on a brightness-discrimination problem until the criterion of 18 correct choices out of 20 successive free trials, with the last 10 correct, was met. Throughout acquisition 20% of each day's 20 trials were forced by locking closed the door to one of the arms. Two of the forced trials were to S+ and two were to S−. For 12 *S*s the positive stimulus was the bright arm (10.0 footcandles [ftc.]), and for 8 *S*s it was the dim (0.1 ftc.). Since the results were comparable for the two conditions, this variable was not maintained as a factor in subsequent statistical analyses.

The "reward interval," during which *S* had access to a full water dipper was approximately 2 sec.; the "stimulus interval," the period during which the discriminanda remained on in the maze after the dipper was retracted, was 8 sec.; the intertrial interval (ITI), spent in near total darkness, was about 60 sec.; finally, "the starting interval," the period signaling the beginning of the next trial, lasted some 5 sec. (For further explanation see D'Amato & Jagoda, 1960.)

Upon reaching acquisition criterion *S*s were assigned to a control group (Group C) and to an overlearning group (Group O) in such a way as to balance the two groups for mean acquisition scores.

Overtraining and reversal.—Overtraining consisted of 240 free trials, distributed 20 per day. Aside from the absence of forced trials overtraining was identical to acquisition training. During reversal, 20 free trials were given daily until *S*s met the same criterion as employed in acquisition; in addition, all *S*s were run a minimum of 160 trials.

Results

Acquisition and overtraining.—Since the control and overtrained groups were always closely equated on acquisition scores, acquisition data will not be presented for each individual experiment. The mean numbers of trials required to reach acquisition criterion ranged between 63.3 and 97.5.

Errors committed during overtraining varied between 1.8% (Exp. VIII) and 5.7% (Exp. IV).

Reversal.—There was no sign of an ORE in the trials-to-criterion data. Group C ($N = 11$) required 145.0 trials on the average to reverse, while for Group O ($N = 9$) the corresponding figure was 153.2. A trend analysis showed the linear components of the reversal curves not to differ significantly, F (1, 18) = .84.

Experiment II

We now embarked upon a series of experiments in an attempt to uncover the factors that prevented the occurrence of an ORE in our experimental situation, the first experiment being motivated by the following hypothesis. Let us suppose that small stimulus changes are usually associated with the transition from acquisition to reversal learning. For example, if *S*s are run in squads, due to the circumstance that *S*s slow up when reversed, the ITI will probably increase during the early portions of reversal learning. This change, as well as others that might occur on the transition from acquisition to reversal, ought to be more readily discriminated by the overtrained *S*s due to their extended experience with the stimulus conditions of acquisition. Consequently, such changes should be more disruptive to overtrained *S*s and result in relatively faster extinction of the formerly correct response. If there were anything to this line of reasoning it could explain our prior failures to obtain the ORE, since the highly controlled conditions of our experimental situation permit of no stimulus changes in reversal other than the shifting of reward from the former S+ to the former S−. In any event, the appropriateness of this hypothesis was explored in the present experiment by imposing a stimulus change—reduction of the stimulus interval—during reversal learning.

Method

In this and subsequent experiments S+ was the bright arm and all trials were run free. The reward interval was 2 sec., and the stimulus interval, 8 sec. There followed a 54-sec. ITI during which the stimuli were extinguished; consequently, *S*'s postchoice exposure to the discriminanda amounted to 10 sec. Ten trials were given on Day 1 of acquisition and 20 per day thereafter. Overtraining consisted of 240 trials beyond the acquisition criterion.

Reversal.—The postchoice exposure to the discriminanda was reduced during reversal from 10 to 4 sec. by reducing the stimulus interval from 8 to 2 sec. To maintain the time between choice and the start of a subsequent trial constant, 6 sec. was added to the ITI, increasing the latter to 60 sec. As usual, the ITI was spent in darkness.

Results

The mean numbers of trials to reversal criterion were 151.4 for Group C ($N = 12$) and 136.9 ($N = 12$) for Group O. The difference, though in the "right" direction, is small and does not approach significance, $t < 1$.

Experiment III

In this experiment we pursued the differential generalization decrement notion further by introducing a more imposing change between acquisition and reversal, at the same time employing larger groups of *S*s.

Method

The only difference between the procedures of the present experiment and those of Exp. II was that, during acquisition and overtraining, the discriminanda remained on for the duration of the ITI. Consequently, during acquisition and overtraining *S* was exposed to the discriminandum in the arm of his choice throughout the trial sequence (a total of 64 sec.), until the beginning of the starting interval of the next trial. During reversal the postchoice duration of the discriminanda was reduced to 4 sec., and *S* spent the next 60 sec. in darkness. All *S*s were run a minimum of 160 reversal trials.

Results

The mean numbers of trials to reversal criterion were 134.8 for Group C ($N = 20$) and 130.1 for Group O ($N = 18$), respectively. The absence of an ORE in the trials-to-criterion data is obvious. The difference between the linear components of the

to the ORE than are our own animals. The difference between the linear components of the reversal curves was small, F (1, 17) = .05.

Experiment VIII

The last experiment to be reported had two purposes. Earlier work had shown that acquisition of a brightness discrimination is consistently impeded in our Y mazes when the discriminanda are permitted to remain during the ITI (D'Amato, 1960). One aim of the present study was to ascertain whether this result extends to pigmented *S*s. Since albinos alone had been used in Exp. I–VII, we continued on through overtraining and reversal on the chance that the pigmented *S*s might show some signs of an ORE.

Method

Subjects.—The *S*s were 28 naive male rats, 80–100 days of age at the start of the study. Drawn from three hybrid litters bred in our laboratory, hooded black and brown *S*s accounted for 26 of the *S*s, and the remaining 2 were albinos.

Acquisition.—The *S*s were trained 20 trials per day on the usual brightness-discrimination problem. For half of the *S*s, randomly chosen, the discriminanda remained on during the ITI (Group A); for the other half (Group B) the discriminanda were extinguished. The two groups were otherwise treated alike. The reward, stimulus, intertrial, and starting intervals were 2, 8, 60, and 5 sec. Upon reaching criterion each *S* was assigned to either a control or an overlearning group, making four groups in all, Groups AC, AO, BC, and BO (N = 7).

Overtraining and reversal.—Overtraining consisted of the usual 240 trials. All *S*s were run a minimum of 140 reversal trials. Perhaps it should be explicitly stated that *S*s of Cond. A remained in the presence of the discriminanda during the ITI throughout overtraining and reversal.

Results

Acquisition.—In verification of earlier results, the 14 *S*s trained with the discriminanda present during the ITI required significantly more trials to learn the brightness discrimination than those in the B condition, the appropriate means being 82.6 and 65.1, respectively, t (26) = 3.00, $p < .01$.

Reversal.—Control Groups AC and BC averaged 170.4 and 134.1 trials, respectively, to meet the reversal criterion. For the corresponding overtrained groups these values were 154.7 and 132.4. Obviously the differences between the control and overtrained groups are quite small; evaluated by a two-way analysis of variance, F (1, 24) = .46. On the other hand, as one might expect from the differences in acquisition, *S*s of Cond. A take significantly longer to reverse than *S*s of Cond. B, F = 5.24, $p < .05$. For the interaction term F = .30. A two-way trend analysis of the first 7 reversal days showed none of the differences between the linear components to be significant.

Discussion

None of the variables investigated in the eight experiments has been effective in producing an ORE, nor is there an indication that a particular combination of these variables would prove successful. The pattern observed repeatedly was a small, nonsignificant, difference between the control and overtrained groups, favoring one group as often as the other, frequently combined with a tendency for the reversal curve of the overtrained group to be slightly steeper than the controls. This pattern is reflected in part by the combined results of the eight experiments, the control (N = 116) and overtrained (N = 114) groups averaging 80.5 and 80.6 trials to acquisition criterion, respectively, and 134.3 and 141.1 trials to reversal criterion.

Obviously we have not explored all possibilities: motive-incentive conditions, task difficulty, as well as a number of other possibly relevant variables were not investigated, and, of course, the same apparatus was used in all experiments.

But we believe that we have studied a sufficient number of variables to demonstrate that the ORE, even in visual-discrimination studies at the rat level, is by no means a general phenomenon, a conclusion bolstered by Erlebacher's (1963) failure to find an ORE with conventional apparatus and hunger-motivated *Ss*. The conclusion seems warranted, therefore, that though the ORE may be a genuine phenomenon, the set of conditions under which it can be expected to occur must be quite specific and cannot be accurately delineated at the present time.

Specificity of results to highly restricted sets of conditions is not, however, an affliction solely of the ORE. Judging from the inconsistencies of results and the failures to replicate so prevalent in contemporary psychological research (despite the reluctance on all sides to publish "negative" results), the generality of the empirical relationships in psychology may be more limited than most of us now realize. If so, the continued accumulation of isolated empirical relationships is not likely to be of any great value in accelerating our understanding of behavior. Perhaps of greater priority is the development of a set of principles which will point up the important dimensions along which the generality (invariance) of interesting empirical relationships ought routinely to be evaluated. The problem is a difficult one, but it must be faced if we are to avoid being buried under a growing pyramid of disjointed and conflicting collections of data, highly restricted to unspecifiable sets of conditions.

REFERENCES

Brookshire, K. H., Warren, J. M., & Ball, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54**, 98–102.

Capaldi, E. J. Overlearning reversal effect in a spatial discrimination task. *Percept. mot. Skills*, 1963, **16**, 335–336.

Clayton, K. N. Reversal performance by rats following overlearning with and without irrelevant stimuli. *J. exp. Psychol.*, 1963, **66**, 255–259.

D'Amato, M. R. Distribution variables in simple discrimination learning in rats. *Canad. J. Psychol.*, 1960, **14**, 216–219.

D'Amato, M. R., & Jagoda, H. Effects of extinction trials on discrimination reversal. *J. exp. Psychol.*, 1960, **59**, 254–260.

D'Amato, M. R., & Jagoda, H. Analysis of the role of overlearning in discrimination learning. *J. exp. Psychol.*, 1961, **61**, 45–50.

D'Amato, M. R., & Jagoda, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, **64**, 117–122.

Erlebacher, A. Reversal learning in rats as a function of percentage of reinforcement and degree of learning. *J. exp. Psychol.*, 1963, **66**, 84–90.

Hill, W. F., & Spear, N. E. A replication of overlearning and reversal in a T maze. *J. exp. Psychol.*, 1963, **65**, 317.

Hill, W. F., Spear, N. E., & Clayton, K. N. T maze reversal after several different overtraining procedures. *J. exp. Psychol.*, 1962, **64**, 533–540.

Komaki, J. Reversal retardation through introducing forced non-rewarding trials in overtraining by white rats. *Ann. anim. Psychol.*, 1962, **12**, 1–10.

Mandler, G. From association to structure. *Psychol. Rev.*, 1962, **69**, 415–427.

Pubols, B. H., Jr. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

Wagner, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

Warren, J. M. Reversal learning by paradise fish (*Macropodus operculares*). *J. comp. physiol. Psychol.*, 1960, **53**, 376–378.

Warren, J. M., Brookshire, K. H., Ball, G. G., & Reynolds, D. V. Reversal learning by white leghorn chicks. *J. comp. physiol. Psychol.*, 1960, **53**, 371–375.

(Received February 12, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 382–386

INFERRED COMPONENTS OF REACTION TIMES AS FUNCTIONS OF FOREPERIOD DURATION

RAYMOND H. HOHLE

Institute of Child Behavior and Development, University of Iowa

A distribution function representing simple reaction-time distributions was derived, assuming RT is the sum of 2 component variables with exponential and normal distributions. 4 Ss each gave 100 RTs to an auditory stimulus following each of 4 foreperiods, under each of 2 conditions: (a) foreperiod constant within sessions but varied over sessions, and (b) foreperiods appearing in a random sequence. The derived distribution function provided completely satisfactory representations of all 32 RT distributions, and the relations of the fitted parameters of this function to foreperiod suggest that the variation of RT as a function of foreperiod is due to variation in the normally distributed component.

Concern with the form of distributions of reaction times (RT) has been focused largely on the fact that the nonnormal character of these distributions tends to preclude application of the more common statistics in studies using RT as a dependent variable. Some attempts have been made recently, however, to relate certain properties of these distributions to hypothetical components of the observed response times.

Restle (1961), for example, observing an almost constant 10 to 1 ratio between the means and standard deviations of 15 distributions of RT (data from Chocolle, 1940), inferred that the gamma distribution function might describe these distributions. This would imply that an observed RT must consist of the sum of a fixed number of independent component times, each of which is a random variable with a common exponential distribution. From the observed ratio of means and standard deviations in Chocolle's data, Restle further inferred that the number of components must be on the order of 100. Unfortunately, as McGill (1963) has pointed out, this interesting result does not correspond to commonly observed data. According to the central-limit theorem, a gamma distribution with a parameter indicating more than, say, 10 or 15 components would be very nearly normal. But observed distributions of simple RTs obtained under constant conditions seem inevitably to be positively skewed.

McGill (1963) observed that the upper portions of plots of several cumulative distributions of simple RT to different auditory stimulus intensities all had the form of exponential distribution functions with similar time constants. This suggested that at least one component of the total RT was exponentially distributed; and since the time constants implied by the curves seemed to be nearly independent of stimulus intensity, McGill assumed that this component is the time required to make the required motor response. Furthermore, since the overall distributions clearly varied with intensity of the stimuli, it was assumed that the RTs contained a nonexponential component which might be interpreted as "decision time."

As part of a model intended to represent the structure of disjunctive RT, Christie and Luce (1956) assumed that an observed response time is composed of an exponentially distrib-

uted choice or decision time plus a variable "base" time or motor response time, the distribution of which was not specified. Both McGill and Christie and Luce thus assume that RT distributions result from convolutions of distributions of two component random variables, one of which has the exponential distribution. Both also hypothesize that the two components represent decision time and a residual latency which is primarily motor response time. But here the agreement ends: the two papers present directly opposite interpretations of the exponential component. McGill (1963) assumed it is the "movement response [p. 329]," whereas Christie and Luce (1956) assumed it to be "decision latency [p. 25]."

Clearly, interpretation of the component processes underlying reaction latencies would be facilitated by knowledge of a distribution function which accurately describes distributions of total observed reaction times. The aim of the present study is to provide evidence for such a function.

First of all, it is known that the distribution of the sum of a large number of random variables is asymptotically normal regardless of the distributions of the component variables, provided no one component contributes disproportionately to the variance of the sum (Cramer, 1946); therefore, if it is assumed that an RT is the sum of time intervals to complete a sequence of separate processes, then the distribution of this sum would necessarily approach normality unless one (or some very small number) of the components contribute substantially more variance than any of the others.

The simplest hypothesis seems to be that an observed RT results from summation of (*a*) a single dominant exponentially distributed variable as proposed by McGill (1963) and by Christie and Luce (1956), and (*b*) a normally distributed variable which may itself be the sum of a number of component variables with similar variances.

To test this hypothesis the probability density and cumulative distribution functions are needed for the variable $t = t_1 + t_2$, where t_1 and t_2 are the exponentially and normally distributed variables. Letting $a = E(t_1)$, $b = E(t_2)$, and $c = \sigma(t_2)$, and assuming t_1 and t_2 are independent, the cumulative distribution function for t was determined to be

$$H(t) \cong \frac{1}{\sqrt{2\pi}} \int_{-\frac{b}{c}}^{\frac{t-b}{c}} e^{-\frac{y^2}{2}}\, dy - \left[\frac{1}{\sqrt{2\pi}} \int_{-\frac{b}{c}-\frac{c}{a}}^{\frac{t-b}{c}-\frac{c}{a}} e^{-\frac{y^2}{2}}\, dy \right] e^{-\frac{t-b}{a}+\frac{c^2}{2a^2}} \qquad [1]$$

Strictly speaking, Equation 1 is an approximation: its derivation entailed the restriction of both t_1 and t_2 to positive values whereas the assumption of normality for t_2 requires that this variable can take any real value. The approximation is very close, however, as long as the ratio of the mean to the standard deviation of the normal variable is of an order of 3 or more.

Equation 1 provided good representations of 12 distributions of RT to onset of low-intensity visual stimuli (.05 and .125 footlambert (ftL.) against a 1.8-ftL. background).[1] In these data the fitted parameter *a* was markedly different for the two intensities, but the fitted *b* was not related to this variable. A tentative interpretation of these results was that RTs do contain an exponentially distributed component and that this

[1] Data presented at the Midwestern Psychological Association meetings at Chicago, May 1963.

component is the decision or "perception" portion of an RT. The remaining portion, the mean of which is represented by the parameter b, was assumed to be the time required for organization and execution of the motor response.

The present study was designed to provide a further test of Equation 1 and to attempt to manipulate the normally distributed component of simple RTs. It has been generally assumed that effects of different foreperiod durations on simple RT are due to changes in *S*'s preparatory set during the prestimulus delay. And evidence seems to indicate that, for clearly suprathreshold stimuli, the preparatory set is primarily muscular (Teichner, 1954). If this is indeed the case, and if the assumption is correct that a normally distributed component of RT includes motor response time, then it should be possible to account for variation in RT due to foreperiod in terms of variation of the b parameter in Equation 1 as a function of foreperiod.

Method

Distributions of simple RT to an auditory stimulus were obtained with four different foreperiods under two conditions: (*a*) constant foreperiod within sessions, and (*b*) random sequences of foreperiods within sessions.

Each trial began with onset of four panel lights surrounding a 4-in. speaker. After a delay (foreperiod) of 1.60, 3.20, 4.65, or 6.13 sec., a 62-db. (re .0002 dynes/cm^2), 1,000-cycle tone appeared from the speaker, which was *S*'s signal to press a microswitch button as quickly as possible. The response terminated both the lights and the tone. There was a constant 1-sec. interval between *S*'s response and the beginning of the next trial. Ambient background noise in the experimental room was 35 db.

Four 14-yr.-old female *S*s each participated in a practice session followed by eight 130-trial sessions. Only the last 100 trials from a session were retained for analysis, the first 30 being treated as warm-up trials. The eight sessions were spread over 10 days with no *S* participating in more than one session on any one day.

During the first four sessions the foreperiods were presented in a random sequence for all *S*s; during the last four the foreperiod was constant within a session but varied over sessions. Two *S*s received the constant foreperiods in descending order, the other two in ascending order. A distribution of 100 responses was thus obtained from each *S* for each foreperiod under each of the two foreperiod conditions.

For each of the 32 distributions, 10 intervals of response times were selected such that 10% of the observations fell in each interval. Parameters in Equation 1 were then determined by successive approximations such that this distribution function assigned as nearly as possible 10% of the probability to each of the 10 intervals. The criterion of closeness of fit was a chi-square statistic based on deviations of observed from expected frequencies in the 10 intervals.

Results and Conclusions

The correspondence between Equation 1 and the four *S*s' distributions of RTs following the shortest foreperiod used is indicated in Fig. 1. The upper four distributions are those obtained with this foreperiod occurring in a random sequence; the lower four were obtained with the foreperiod constant. That this generally excellent correspondence was uniform over the 32 sets of data is indicated by the distribution of the goodness-of-fit statistic: Under the null hypothesis—i.e., the hypothesis of perfect fit of Equation 1 except for random sampling error—this statistic would be distributed as a chi-square variable with $df = 6$. Thus if a perfect fit is assumed, the expected value of the statistic for any one distribution would be 6, and the expected value of the sum of the 32 (independent) statistics would be $(32)(6) = 192$. Obtained values of the 32 statistics ranged from .65 to 11.85, and their sum was 172.17. The median value was 4.92 which may be compared to an expected median of 5.35 on the basis of random sampling.

Figure 2 shows the obtained relationships of mean RT and its inferred components to foreperiod duration

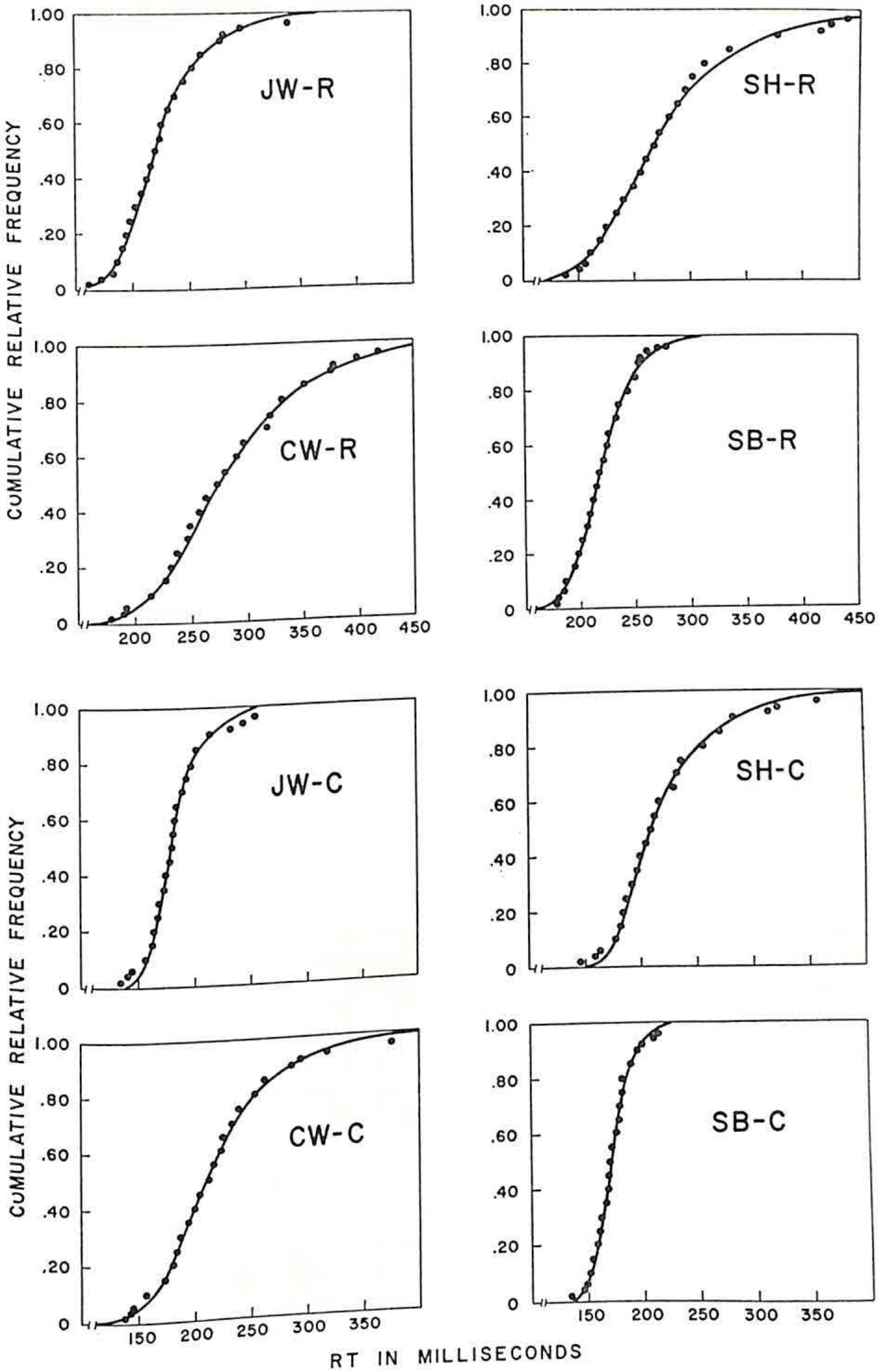


FIG. 1. Equation 1 fitted to the four *S*s' cumulative distributions of RTs to a 62-db., 1,000-cycle tone with a foreperiod of 1.60 sec. (The plotted points are the second, fourth, sixth, ninety-second, ninety-fourth, and ninety-sixth percentiles, and every fifth percentile from the tenth to the ninetieth.)

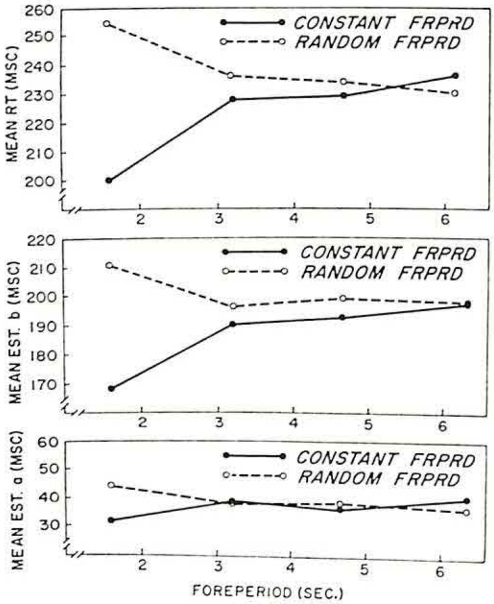


FIG. 2. Relations of mean RT and its inferred components to foreperiod duration. (The parameter a is the mean of a hypothetical exponentially distributed component; b is the mean of a hypothetical normally distributed component.)

under the two conditions. The mean RT curves are very similar to those reported in other studies (e.g., Karlin, 1959); the shortest foreperiod appears to have been disruptive when it occurred in a random sequence including longer foreperiods, but resulted in the shortest RTs when it was constant from trial to trial. Figure 2 suggests that most of this effect can be accounted for as differential variation of b, the normal component, with foreperiod under the two conditions.

A Foreperiod × Constant-vs.-Random × *S*s analysis of variance for mean RT and for each of the fitted parameters indicated that (*a*) the Foreperiod × Constant vs. Random interaction effect on mean RT was significant at the .05 level, $F\,(3, 9) = 6.61$; (*b*) the corresponding obtained F for the b parameter was 3.81, which is very close to the critical value (3.86) required for significance at the .05 level; while (*c*) this interaction effect on the parameter a did not approach significance ($F = 1.46$); and (*d*) although the parameter c, interpreted as the standard deviation of the normally distributed component, showed considerable variation among *S*s, its variations due to Foreperiod and Constant vs. Random conditions were not statistically significant. The range of estimated c's over the 32 sets of data was from 10 to 42 msec.

The distributions of RTs to low-intensity visual stimuli mentioned earlier and the auditory RT data reported here were thus both consistent with the hypothesis that an observed distribution of simple RTs may be described by a convolution of a normal and an exponential distribution. Moreover, the apparent differential dependence of the means of these inferred component distributions on stimulus intensity and foreperiod duration strongly suggests that the component distributions are in fact associated with times to complete different processes underlying observed RTs.

REFERENCES

CHOCOLLE, R. Variation des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *Annee psychol.*, 1940, **41**, 5–124.

CHRISTIE, L. S., & LUCE, R. D. Decision structure and time relations in simple choice behavior. *Bull. math. Biophys.*, 1956, **18**, 89–112.

CRAMER, H. *Mathematical methods of statistics.* Princeton: Princeton Univer. Press, 1946.

KARLIN, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58**, 185–191.

MCGILL, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* New York: Wiley, 1963.

RESTLE, F. *Psychology of judgment and choice: A theoretical essay.* New York: Wiley, 1961.

TEICHNER, W. H. Recent studies in simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–149.

(Received February 13, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 387–392

FIRST-LIST RETENTION AS A FUNCTION OF LIST DIFFERENTIATION AND SECOND-LIST MASSED AND DISTRIBUTED PRACTICE [1]

JOHN P. HOUSTON [2] AND JAMES H. REYNOLDS [3]

University of California, Berkeley

The effects of list differentiation on retention following massed and distributed practice were evaluated by measuring 1st-list retention, or RI, in an A-B, A-C situation. A-C learning was either by massed or distributed practice. A-B and A-C learning were either by the same or different practice methods. 1st-list recall was better when A-B and A-C were learned by different procedures than when they were learned by the same procedures. 1st-list recall following distributed A-C learning did not differ from 1st-list recall following massed A-C learning.

Underwood, Keppel, and Schulz (1962) have attributed the facilitative effects of distributed practice (DP) to a successive extinction-recovery process whereby interfering associations are more effectively and permanently extinguished during distributed learning than during massed learning. In discussing their finding that A-E recall in an A-B, A-C, A-D, A-E situation was greater following distributed A-E learning than following massed A-E learning, these authors suggested that the successive A-E distributed practice intervals allowed for greater recovery and therefore more effective extinction of the associations established between the context stimuli and the responses of the first three lists during A-E learning. Thus, at the time of A-E recall 24 hr. after A-E learning, proactive inhibition (PI) associated with the interfering context associations would have been weaker following distributed A-E learning than following massed A-E learning, allowing the facilitative effect to appear.

There is another process, involving list differentiation, which could contribute to the facilitative effects of DP. In the Underwood, Keppel, and Schulz experiment all *S*s learned the first three lists by massed practice (MP). A DP group then learned the A-E list by DP while an MP group learned A-E by MP. Thus, the MP group learned all lists by MP while the DP group learned the A-E list by a procedure which differed from the conditions under which the first three lists were learned. The superior A-E recall following DP may have been due, in part, to the DP *S*s' ability to differentiate the lists on the basis of the difference between the practice methods employed in learning the lists. The MP *S*s, because they learned all lists by MP, would be subject to a greater degree of list confusion than would the DP *S*s. The implication is that the superior A-E recall demonstrated by the DP group may have been a function of both differential extinction-recovery and list differentiation.

The present experiment was designed to evaluate the list differentiation interpretation of the DP effect by

[1] This study was carried out during the tenures of the authors' United States Public Health Service postdoctoral fellowships at the Institute of Human Learning, University of California, Berkeley.

[2] Now at University of California, Los Angeles.

[3] Now at Colgate University.

measuring retention of first-list associations 1 min. after A-B, A-C learning. This retroactive inhibition (RI) paradigm was employed because it was thought that RI would be a more direct measure of the effects of DP and MP on previously learned associations than would a PI measure as has been used in previous studies. The main effects of the 2 × 2 factorial design were massed vs. distributed A-C practice and same vs. different practice procedures employed in learning the A-B and A-C lists. The four main conditions were massed A-B learning followed by massed A-C learning (MP-MP), massed A-B followed by distributed A-C (MP-DP), distributed A-B followed by distributed A-C (DP-DP), and distributed A-B followed by massed A-C (DP-MP). The list-differentiation hypothesis predicts that first-list retention will be greater in the two paradigms involving different A-B and A-C practice procedures (MP-DP, DP-MP) than in the two conditions involving the same A-B and A-C practice procedures (MP-MP, DP-DP). A prediction with respect to the extinction-recovery hypothesis is more difficult because there has been no clear statement as to whether DP is expected to result in a lower initial level of extinction as well as slower recovery. If it is assumed that the extinction-recovery process implies a greater loss of Context B associations during DP than during MP then it may be predicted that retention of first-list associations in the two distributed A-C learning conditions (MP-DP, DP-DP) will be less than the first-list retention in the two massed A-C conditions (MP-MP, DP-MP).

Method

General.—All *S*s learned two successive paired-associate lists conforming to an A-B, A-C paradigm. Depending upon the list and the condition, learning was either by MP or DP. The *S*s were given nine anticipation trials on both the A-B and the A-C lists. Distributed intervals were filled with a digit-cancellation task. Following A-C learning all *S*s were given a recall trial and four additional relearning trials on the A-B list to evaluate RI. It was assumed that differences in the Context B associations would be reflected in the recall of specific A-B associations because the context association is a prerequisite of the specific association; i.e., a response must be available if it is to be associated with a specific stimulus.

The general design called for four practice sequences (MP-MP, DP-DP, MP-DP, DP-MP). The variations in second-list practice procedures presented a problem in that 10 distributed A-C trials take longer than 10 massed A-C trials, making the interval between the end of A-B learning and recall longer for the MP-DP and the DP-DP conditions than for the MP-MP and DP-MP conditions. This problem was eliminated, and the time between A-B learning and recall equated for all conditions, by having the MP-MP and DP-MP groups cancel digits for 9 min. (the amount of time the DP-DP and MP-DP groups canceled digits during distributed A-C learning) either just before or after A-C learning. This was accomplished by splitting the MP-MP and DP-MP conditions into two groups each. The MP-C-MP and the DP-C-MP groups canceled digits between A-B and A-C learning while the MP-MP-C and the DP-MP-C groups canceled digits between A-C learning and recall. Thus the complete design, including the groups developed to account for the differences in time required to learn by MP and DP, required six groups (MP-DP, DP-DP, DP-C-MP, DP-MP-C, MP-C-MP, MP-MP-C).

Subjects.—The *S*s were 90 University of California undergraduates, 15 per group, whose participation was in fulfillment of a course requirement. The first 45 *S*s received MP on the A-B list and were randomly assigned to one of the three A-C conditions (DP, MP-C, C-MP) as they appeared for the experiment. The remaining *S*s received DP on the A-B list and were randomly assigned to the A-C conditions.

Lists.—The A-B list for all *S*s was List 1 in Table 1 of the Underwood and Schulz (1961) experiment. The stimulus terms were nonsense syllables and the response terms two-syllable adjectives. The A-B and A-C stimulus terms were identical. The A-C

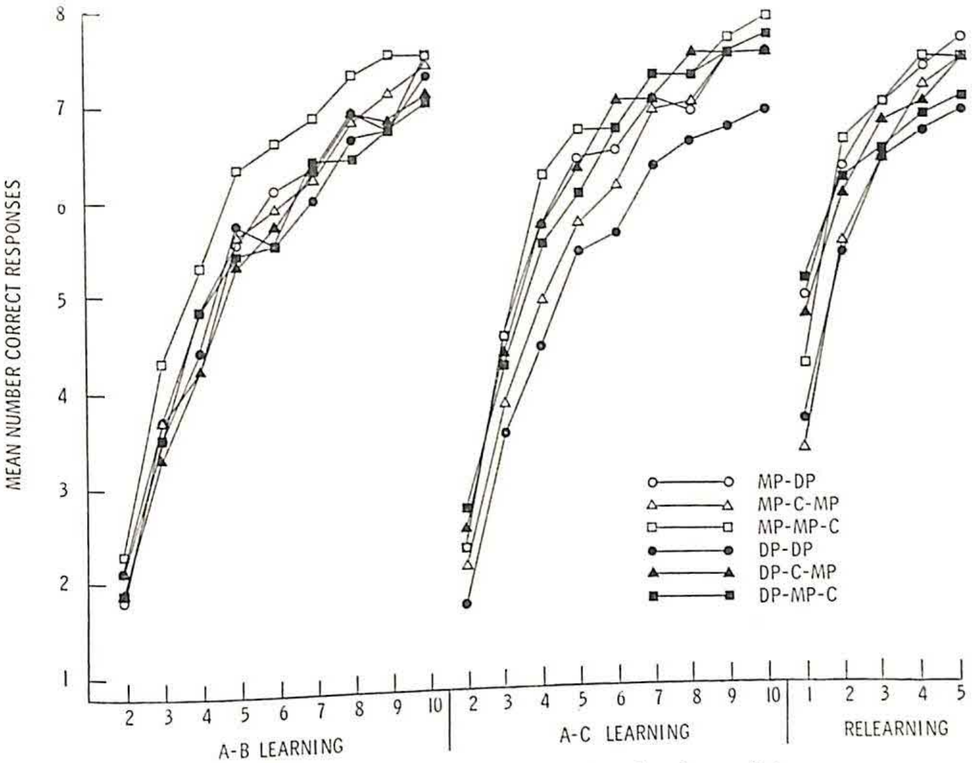


FIG. 1. Learning and relearning curves for the six conditions.

response terms were the adjectives from List 2 of Table 1 (Underwood & Schulz, 1961).

Procedure.—Lists were presented at a 2:2-sec. rate on a Stowe drum. Both A-B and A-C lists were presented for nine anticipation trials. Four different orders of the items in each list were used to avoid serial learning. Each order was used equally often as the starting order. The MP intertrial interval was 4 sec. and the DP interval 1 min. The DP intervals were filled with a digit-cancellation task. Approximately 30 sec. separated the learning of the two lists except in those cases where an additional 9 min. of digit cancellation was introduced before or after massed A-C practice. The recall trial, administered 1 min. after the end of the A-C time period, was a 2:2-sec. anticipation trial. Relearning was then carried for an additional four trials. Prior to recall and relearning *Ss* were instructed to recall as many of the first-list responses as possible on the recall trial.

RESULTS

Acquisition.—The acquisition and relearning curves for all six experimental treatments are shown in Fig. 1. The mean numbers of correct responses over the nine A-B anticipation trials for the MP-DP, MP-MP-C, and MP-C-MP conditions were 49.47, 53.87, and 49.07, respectively. Mean correct first-list responses for the DP-DP, DP-MP-C, and DP-C-MP groups were 46.87, 47.60, and 47.40, respectively. In order to compare first-list performance of groups which later received the MP-C, C-MP, or DP conditions on the second list, and to compare A-B performance of the groups receiving the same or different List 2 practice conditions, the correct responses on the first list were evaluated in a 3 × 2 analysis of variance. The obtained *F* values were less than 1.00 for the two levels of differentiation, the three levels of second-list practice, and the interaction, implying that the groups comprising the various experimental conditions did not differ

significantly in terms of first-list learning.

The mean numbers of correct responses over the nine A-C anticipation trials were 54.33, 56.67, 51.67, 47.13, 55.73, and 55.33 for the MP-DP, MP-MP-C, MP-C-MP, DP-DP, DP-MP-C, and DP-C-MP conditions, respectively. These second-list data were also compared in an analysis of variance based on the two levels of differentiation and the three levels of second-list practice. Although there was a tendency for groups receiving different practice conditions on List 2 to learn the second list better than groups receiving the same practice conditions on both lists, as might be predicted from the differentiation hypothesis, the difference was not significant, $F(1, 84) = 1.60$; $p > .05$. The second-list performance for groups receiving the DP practice condition on List 2 was somewhat below the groups receiving both the MP-C and the C-MP treatments, but this difference was also not significant, $F(2, 84) = 1.48$; $p > .05$. The latter result is consistent with the Underwood, Keppel, and Schulz (1962) finding that DP had a slight but nonsignificant effect upon A-E learning in the A-B, A-C, A-D, A-E situation. The interaction was not significant, $F(2, 84) = 1.10$, $p > .05$. The results of the analysis indicate that the experimental conditions had no significant effects upon acquisition of the second list.

A-B responses appeared as intrusions in the A-C learning of the DP-DP, MP-MP-C, MP-C-MP, DP-MP-C, and DP-C-MP conditions six, five, six, four, and five times, respectively. The MP-DP group gave 17 intrusions but 8 of these were produced by one *S*. The low numbers of intrusions, plus the fact that one *S*'s responses account for the major portion of the between-conditions variance, suggest that little should be made of these intrusion data.

Recall and relearning.—Table 1 contains the recall and relearning means and *SD*s for the six cells of the factorial design. The differentiation hypothesis predicts greater first-list retention for groups receiving differing treatments on the A-B and A-C lists, regardless of whether A-C was learned by DP or MP, because these groups had an opportunity to differentiate the two lists on the basis of the different practice procedures employed during learning. The present interpretation of the extinction-recovery hypothesis implies poorer first-list retention (greater RI) following distributed A-C learning, regardless of the A-B practice conditions. To test these predictions the recall and relearning measures contained in Table 1 were compared in 3 × 2 factorial analyses of variance containing the three levels of second-list practice (MP-C, C-MP, DP) and the two levels of list differentiation (same, different).

Analysis of the recall scores indicated no significant difference among the three A-C practice conditions,

TABLE 1

MEAN NUMBERS OF A-B RESPONSES CORRECTLY REPRODUCED IN RECALL AND RELEARNING

A-C Practice Cond.	A-B, A-C Differentiation Cond.			
	Same		Different	
	M	*SD*	*M*	*SD*
MP-C				
Recall	4.20	1.94	5.13	2.28
Relearning	32.80	5.52	31.80	5.14
C-MP				
Recall	3.27	2.17	4.67	1.89
Relearning	29.80	6.24	32.00	7.16
DP				
Recall	3.60	2.53	4.93	2.02
Relearning	29.07	7.47	33.40	5.64

$F\ (2, 84) < 1.00$. The list-differentiation effect was significant, $F\ (1, 84) = 6.80$, $p < .05$, with groups receiving different types of practice on the first and second lists showing better A-B recall than those receiving the same practice procedure on both lists. The interaction was not significant, $F\ (2, 84) < 1.00$.

The *S*s receiving the same practice conditions on both lists produced 17 A-C intrusions in A-B recall while *S*s receiving different practice conditions gave 7 intrusions. While the numbers are small, and little importance should be attached to the result, these data do conform to the list-differentiation expectation that there will be fewer A-C intrusions in A-B recall when A-B and A-C practice conditions are different than when they are the same.

The analysis of variance performed on the relearning data showed no significant differences for either the list differentiation, $F\ (1, 84) < 1.00$, or the second-list practice variables, $F\ (2, 84) < 1.00$. The interaction was not significant, $F\ (2, 84) < 1.00$. The relearning curves plotted in Fig. 1 illustrate that the list-differentiation effect evident on the recall trial disappeared quickly during succeeding relearning trials.

The numbers of A-C intrusions in A-B relearning for the same and different conditions during the five recall and relearning trials were 39 and 21, respectively. While the numbers were again very small they did conform to the expectation, derived from the list-differentiation interpretation, that the same practice procedures would lead to a greater number of intrusions than would different practice procedures.

Discussion

The principal findings of the experiment were as follows. When *S*s learned two lists in an A-B, A-C situation, with either MP or DP on *both* the first and the second list, recall of the A-B list following A-C learning was poorer than if *S*s learned the first and second lists by different practice procedures (either DP of A-B and MP of A-C or vice versa). When A-B practice conditions were counterbalanced, A-B recall following distributed A-C learning did not differ from A-B recall following massed A-C learning. The implications of these findings for the extinction-recovery and the list-differentiation hypotheses presented in the introduction will be discussed separately.

According to the present interpretation of the extinction-recovery hypothesis, B recall should be poorer following distributed A-C learning than following massed A-C learning, regardless of the conditions of prior practice. The lack of a DP effect may be interpreted in several ways. First, it could conceivably reflect the fact that DP may not, by itself, lead to more effective extinction of prior associations in a transfer situation. However, this interpretation cannot be seriously considered on the basis of this one study which failed to find DP effects with an RI measure. A more likely interpretation is that MP and DP lead to different rates of B recovery over extended periods of time but not, as has been assumed in this paper, to different levels of B strength immediately following A-C learning. According to this interpretation, which is perhaps the interpretation intended by Underwood, Keppel, and Schulz, both MP and DP extinguish first-list responses to the same level but result in different rates of recovery. Thus, the present experiment, measuring retention 1 min. after A-C learning, may not have allowed sufficient time for differential recovery to occur and was, therefore, an inadequate test of the hypothesis. A more appropriate test would involve first-list retention after a 24-hr. retention interval with list differentiation controlled. A third explanation of the failure to find a significant second-list practice effect would hold that a more appropriate test of the availability

under eight different exposure durations and eight luminances, by the method of constant stimuli. Exposure durations varied from 4 msec. to 512 msec., in $\log_2$ steps. Luminance was varied in $\log_2$ steps from .09 mL. to 11.84 mL. The fixation field luminance was always equal to that of the stimulus field. Within a session each of the eight luminances was constant for a block of 16 exposures (every exposure duration for each pattern), for a total of 128 exposures for each session. The order of luminance blocks was randomized, with the restriction that the initial block was always in the brighter half of the range and that successive blocks were within two steps of each other. Pattern-exposure-duration combinations were randomized within each block, with each combination used only once within a block.

Approximately .5 sec. before each exposure a warning signal was given with a manually operated buzzer, and *S* focused on the space between two horizontal edge markers on the fixation field. To respond, *S* was asked to hit, as quickly as possible, one of two microswitch keys located on either side of a finger rest. For three of the *S*s, the key on the right side was associated with the evenly spaced designs; for two of the *S*s this key was associated with the paired designs. The designs were not described verbally, either by *S* or by *E*, until the end of the study. If *S* saw absolutely no indication of a design during the exposure (i.e., not even a blurred line), he was instructed to hit either key, and then say the word "nothing." There was an interval of approximately 11 sec. between exposures within blocks of exposures and 45 sec. between each block, when intensity filters were changed.

Subjects.—The *S*s were five male college seniors who were paid for their participation. None had previous experience in perception experiments.

Results

Figure 1a shows the mean RT for each pattern as a function of exposure duration, averaged across all luminances. Figure 1b shows the mean RT for each pattern at every luminance, averaged across all exposure durations. Each point in Fig. 1a and 1b is based on correct responses from a total of 450 exposures. Only the RTs of correct responses were selected in the belief that these would most directly reflect differences in discrimination difficulty of these patterns.

Figures 1a and 1b show that, in general, RTs become faster with increases in duration and luminance, but exceptions occur at the longest duration and brightest luminance where there is some tendency for RT to become slower again.

Figures 1a and 1b also indicate an RT ordering of patterns which is similar to the ordering obtained for accuracy of identification (Kaswan & Young, 1963). With both accuracy and RT measures, the obtained ordering from lowest to highest accuracy and fastest to slowest RT was $Pr_1 < Pr_2 < Pr_3$ and $Ev_1 < Ev_2 < Ev_3$. Also, as for accuracy, RT to paired patterns tended to be slower than the RTs to comparable Ev patterns, except for the Ev_3-Pr_3 set, where RT to the two patterns was about the same, probably because these patterns were easily discriminated.

While RT reflects the order in which these patterns were accurately discriminated, variations in luminance appear to affect RT and accuracy differently. For accuracy, it was found that discrimination of the patterns depended largely on exposure duration and that variation of luminance had little effect on the level of accuracy. Figure 1b shows, however, that RT was consistently and substantially affected by variations in luminance.

While Fig. 1a and 1b show the "main effects" of exposure duration and luminance on RT, Fig. 2 shows how these variables interact in their effect on RT. The parts of Fig. 2 labeled a through h contain the results of the experiment described above. The parts of the figure labeled a′ through h′ report the results of the supplementary study which will be described later.

Each point in Fig. 2a through 2h is a mean RT based on 340 responses, the result of combining correct, incorrect, and "nothing" responses to

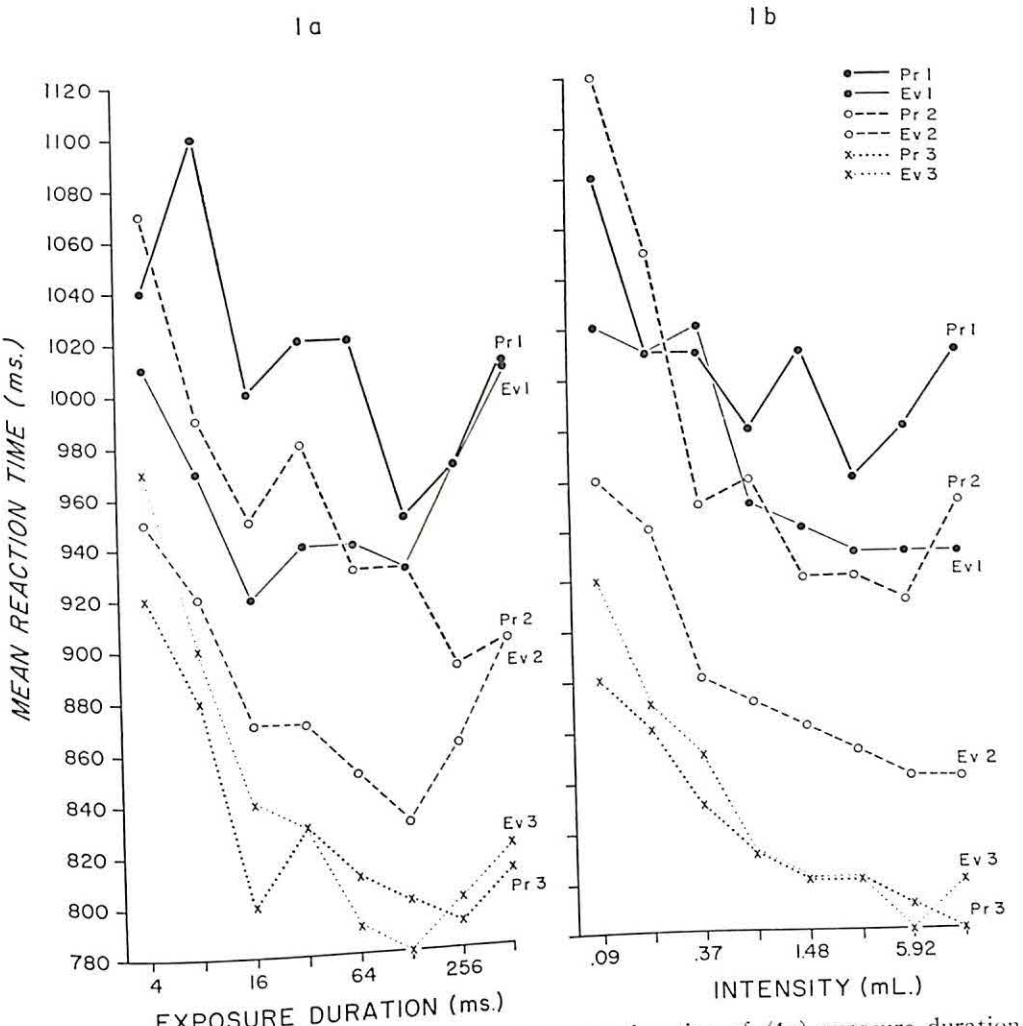


FIG. 1. Mean reaction time for each pattern as a function of (1a) exposure duration averaged over all luminance values for the main experiment and as a function of (1b) luminance averaged over all exposure-duration values.

all six stimuli by the five *S*s. There was close similarity in the *form* of the curves for each pattern and each type of response, although incorrect and "nothing" responses were slower and more variable than correct responses. Figure 2 is therefore quite representative of the results. Vertical bars at representative points in each curve, indicating the standard error at each of these means, show that variability is small relative to the range of RT effects obtained as a function of luminance and duration. Only mean RTs from the sixth through the forty-first session are included. The first five sessions were eliminated because RT decreased over these sessions. During the last two sessions, RT became very variable, perhaps because *S*s anticipated the end of the experiment.

Each of the solid curves in Fig. 2 shows the effect of exposure duration on RT under a different luminance parameter. The exposure-duration curve for the brightest parameter (11.84 mL.) is shown as the solid line in Fig. 2a. Luminance values then become successively dimmer, with the solid curve under the dimmest parameter (.09 mL.) shown in Fig. 2h. Solid curves are to be read against the

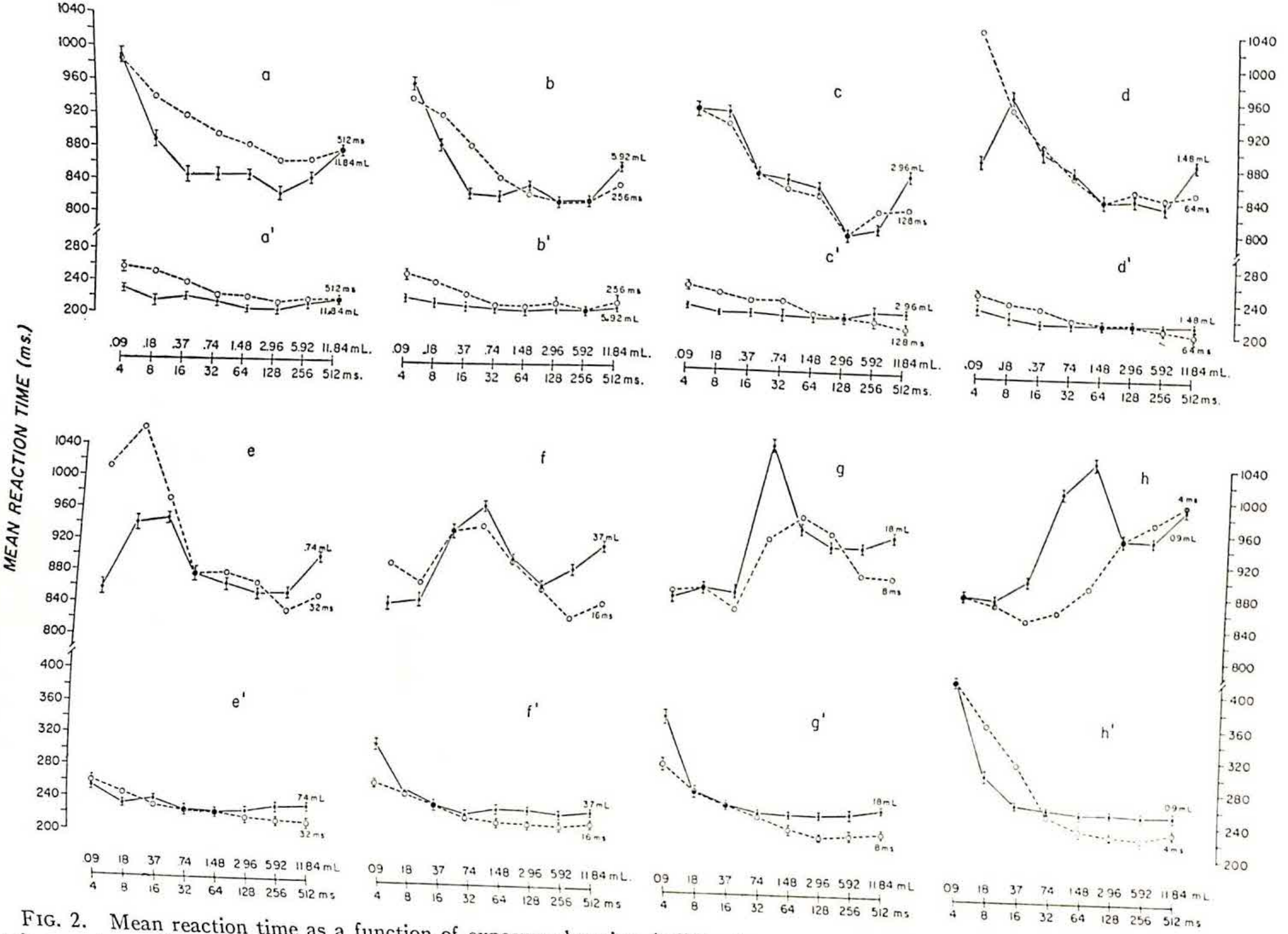


FIG. 2. Mean reaction time as a function of exposure duration (solidly drawn curves) and luminance (dotted curves) for the main experiment (a through h) and the supplementary experiment (a′ through h′). (Vertical bars through data points show the *SE* of the mean [see text].)

exposure-duration scale shown on the abscissa (labeled "ms.") in each part of Fig. 2. It is apparent that the exposure-duration-RT function changes with the level of luminance. As luminance parameters become dimmer (solid lines, Fig. 2a to 2h), the slowest RT occurs at increasingly longer exposure durations. Thus, the slowest RT occurred at the 4-msec. exposure under the 11.84-mL. luminance parameter (Fig. 2a), at the 8-msec. exposure under the 1.48-mL. parameter (Fig. 2d), at about 16 msec. under the .74-mL. parameter (Fig. 2e), at 32 msec. under the .18-mL. parameter (Fig. 2g), and at 64 msec. under the .09-mL. parameter (Fig. 2h). This shift in the exposure duration for which the slowest RT occurred leads to a gradual inversion of the slope in the RT-exposure-duration function. At one extreme, under maximum luminance (11.84 mL., Fig. 2a) RT became *faster* between 4 and 16 msec. At the other extreme, under minimum luminance (.09 mL., Fig. 2h) RT became *slower* between 4 and 64 msec. Thus the slope of the RT function in Fig. 2h is in the opposite direction to the slopes shown in Fig. 2a, 2b, 2c, and 2d.

The curves drawn with dashed lines in Fig. 2 show the effect of luminance on RT at each exposure-duration parameter. For example, the dashed curve in Fig. 2a shows the effect of luminance on RT for the 512-msec. parameter. Exposure-duration parameters (shown next to each curve) then become successively briefer, with the briefest (4 msec.) for the dashed line in Fig. 2h. These curves are to be read against the luminance scale (labeled "mL.") shown below the exposure-duration scale in each part of Fig. 2.

The scales are arranged so that points directly above and below each other on the dashed and solid curves represent equal luminance-exposure-duration products. There is a tendency toward superimposition of the dashed and solid curves, indicating that changes in luminance and exposure duration generally have a similar effect on RT under these conditions. Figures 2d and 2e indicate, however, that the shift toward a reversal in the direction of the RT slope appears to occur a little later in the luminance curves (dashed lines) than in the exposure-duration curves (solid lines).

Since the shift in the direction of the slope of these curves occurred gradually with decreasing parameter values for both luminance and exposure duration, these results appear to constitute a reliable finding.

Figures 2a through 2h show that the present results differ in two major ways from RT findings in flash detection. One difference is that at low luminances (e.g., solid line, Fig. 2h) RT becomes *slower* as exposure duration increases. The same effect was obtained for increases in luminance under brief exposure durations (e.g., dashed line, Fig. 2h). At dim and brief parameter values previous studies found their most substantial RT variations in the *opposite* direction from those reported here. That is, RT tended to become *faster* with increasing exposure duration or increasing luminance, given, respectively, dim or brief parameter values.

A second difference between these results and previous findings relates to the difference in the size of the range of exposure durations and luminances affecting RT. As noted, Raab and Fehrer (1962) found that even at their dimmest luminance parameter (.3 ftL.), exposure durations longer than 20 or 25 msec. had little effect on RT, whereas the range of effective luminance values was much wider (up to 3,000 ftL.). Figures 2a through 2h show, on the other hand, that expo-

changes at bright luminances, since both their data, and ours, indicate that, as luminance levels increase, greater increments in absolute brightness seem necessary to produce equivalent RT changes.

Discussion

The substantial similarity of the Raab and Fehrer findings and the results of our supplementary study indicate that the luminance conditions used (i.e., flash or constant illumination of fixation and exposure field) make little difference in how luminance and exposure duration affect RT. The similarity in the results of these studies is especially striking in view of the differences in apparatus and procedures used.

The results further suggest that the range of exposure durations affecting RT in a detection task is similar to the range affecting detection thresholds. Thus, in the supplementary study, exposure durations longer than 16 msec. did not substantially affect RT (Fig. 2a′ through 2h′). Similarly, in a previous study, probability of detecting a single large dot under dim luminance (.09 mL.) was most affected at exposures briefer than 16 msec. (Kaswan & Young, 1963).

The Raab et al. and our findings also indicate that RT in detection tasks is affected by a much wider range of luminance than exposure-duration values. This suggests that the effect of *luminance* in effecting minimum RT is different from the effect of luminance on "high criterion" detection thresholds (near 100% detection). The latter appears at least grossly determined by a reciprocal luminance-duration relationship (Bunsen-Roscoe "law"), whereas the *minimum* RT cannot be reached by increasing exposure duration to compensate for lower luminance (compare, for example, RTs to 300 ftL. and 3,000 ftL. flashes in the Raab papers).

The agreement with the Raab and Fehrer findings also indicates that the puzzling results of our main study—the reversal in the direction of RT slopes at dim and brief parameters and the extended effect of exposure duration and luminance on RT—were probably due to the complexity of the discrimination task and do not depend on the way in which luminance and exposure duration were varied.

In a previous study the authors suggested that luminance and exposure duration function differently in determining probability of detection and accuracy of pattern discrimination (Kaswan & Young, 1963). The present findings suggest a similar conclusion with regard to reaction time, since luminance and exposure duration clearly generate different RT functions in detection, as compared to pattern discrimination (compare Fig. 2a through 2h with Fig. 2a′ through 2h′). The distinction between discrimination and detection may thus have more than semantic merit and warrants further investigation.

References

Baumgardt, E., & Segal, J. Facilitation et inhibition parametres de la fonction visuelle. *Annee psychol.*, 1942, **43**, 54–102.

Froeberg, H. The relation between magnitude of the stimulus and the time of reaction. *Arch. Psychol., N. Y.*, 1907, **25**, No. 8.

Kaswan, J., & Young, S. Stimulus exposure time, brightness, and spatial factors as determinants of visual perception. *J. exp. Psychol.*, 1963, **65**, 113–123.

Raab, D., & Fehrer, E. The effect of stimulus duration and luminance on visual reaction time. *J. exp. Psychol.*, 1962, **64**, 326–327.

Raab, D., Fehrer, E., & Hershenson, M. Visual reaction time and the Broca-Sulzer phenomenon. *J. exp. Psychol.*, 1961, **61**, 193–199.

Teichner, W. H. Recent studies of simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–149.

Wells, G. R. The influence of stimulus duration on reaction time. *Psychol. Monogr.*, 1913, **15**(5, Whole No. 66).

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received February 17, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 401–407

SIGNAL UNCERTAINTY AND SLEEP LOSS [1]

HAROLD L. WILLIAMS,[2] OMETTA F. KEARNEY, AND ARDIE LUBIN [3]

Walter Reed Army Institute of Research

During a 3–5 day base-line period, 2 days of sleep loss, and 3 days of recovery, 52 Ss performed 3 visual vigilance tasks, of 10 min. each ranging in signal uncertainty from complete redundancy to .84 bit per second. The major effect of uncertainty was to cause errors of omission which increased with sleep loss. The interaction between signal uncertainty and sleep loss was significant. Task duration (of 10 min.) caused no impairment during the base-line and recovery phases, but during sleep loss, errors of omission rose sharply on the last 3 min. of each task. There was no significant interaction between signal uncertainty and task duration. Decrement was considerably greater for Ss working alone than for Ss working in a group. Oral temperature had no consistent relation to errors of omission or to sleep loss.

During sleep loss, impairment of performance seems to be related to the uncertainty of the task, but the nature of the relationship is not clear. Wilkinson (1958) suggested that performance decrement during sleep deprivation increased as task predictability increased. Thus, "complex" tasks involving multiple displays, high information load, or a variety of responses should maintain a sleep-deprived operator in an "aroused" condition. In general Wilkinson showed, as had several other investigators (Katz & Landis, 1935; Robinson & Robinson, 1922), that "complex" tasks were not very sensitive to sleep loss.

A similar explanation has been given for the results found with complex and simple tasks used in studies of signal detection. When simple displays were used, most studies found a decline in percentage of signals detected over time (McGrath, Harabedia, & Buckner, 1959). Investigators who used more complex tasks found no decrement in signal detection (Adams, Stetson, & Humes, 1961; Broadbent, 1950).

Preliminary studies in our laboratory failed to confirm the suggestion that increased task uncertainty raises the efficiency of the sleep-deprived *S*. Performance on a vigilance task designed with high signal uncertainty was greatly impaired during sleep loss, although a redundant task showed no deterioration. The present studies were designed to confirm these results, and to test for interaction between the effect of signal uncertainty and the effect of sleep loss. The experiments also permitted examination of the effects of task duration, group testing, and reduced body temperature.

METHOD

Subjects.—Fifty-two United States Army enlisted volunteers (mean age, 24 yr.; range, 18–45 yr.) took part in the experiment. Intelligence scores obtained by averaging the Vocabulary and Arithmetical Reasoning scores of the Army Classification Battery (Montague, Williams, Lubin, & Gieseking, 1957) ranged from one *SD* ($\sigma = 20$) below the

[1] The authors wish to thank Allen M. Granda and Henry C. Morlock, Jr., for reading the manuscript and making valuable suggestions.

[2] Now at the University of Oklahoma School of Medicine.

[3] Now at the United States Navy Neuropsychiatric Research Unit, San Diego, California.

Army mean of 100 to two *SD*s above; the mean was 105.

Four *S*s were quartered at a time on a ward. The first (reception) day was devoted to familiarization with the schedule and the tests. After 3–5 days of base-line practice on a regular sleep schedule, *S*s had 64 hr. (2 nights) of sleep loss followed by 3 days of recovery testing. During the sleep-loss phase, *S*s were kept awake constantly by *O*s on a 24-hr. schedule. On base-line and recovery days, all *S*s went to sleep between 2200 and 2300 and were awakened at 0600. The tasks were given between 1330 and 1500. During the remaining hours, *S*s were occupied with other activities.

Twenty-four *S*s (Group 1) were tested individually. To examine the effects of working in a group, the remaining 28 (Group 2) were tested in groups of four. For this latter group, oral temperatures were obtained at the beginning of each test session.

Apparatus and procedure.—The vigilance apparatus[4] consisted of a display panel of five lights (red, yellow, green, blue, and white) arranged in a pentagon with sides of $2\frac{1}{2}$ in. The diameter of each light was 1 in., and the center of the pentagon was 60 in. above the floor. The *S* sat at a table at a viewing distance of 8 ft. The response key was a microswitch on a foam rubber base, fastened to the table in a position convenient to *S*'s preferred hand. During group testing sessions, the four *S*s viewed the same display from individual booths arranged side by side.

On the reception day *S*s were taught to press the response key each time the red light came on and to refrain from pressing for all other lights. Correct responses were recorded on counters.

Three tasks differing in predictability were used: a standard task (S), a redundant task (R), and an uncertain task (U). Each task was programed with a five-channel punched tape, and a Western Union tapereader. The tape for each task consisted of 600 stimuli which ran at a rate of one light per second for 10 min. Eight lights out of 30 were signals (red). The tapes were run in the order S, U, R, with approximately 30-sec. rests between tasks.

The S task was modeled closely on the Continuous Performance Test (CPT) described by Rosvold, Mirsky, Sarason, Bransome, and Beck (1956). The same proportion and spacing of signals as well as the same interstimulus interval were used. The eight red signals out of 30 stimulus lights appeared in Positions 6, 9, 11, 14, 18, 23, 27, and 30. Each *S* saw this fixed sequence 20 times during a test session (10 min.). The same tape was used each day.

The U tape was randomized with respect to the 5 lights, subject only to the restriction that 8 lights of each set of 30 were red signal lights. A different 10-min. U tape was used each day.

The R tape consisted of 30 lights in a fixed sequence. The 8 red lights were together. The *S* immediately learned to press the key eight times in a row after 22 nonred lights had been presented. Each *S* saw the fixed R sequence 20 times (10 min.) on each day of testing. The same R tape was used every day.

Stimulus uncertainty for the S task was initially 0.84 bit per stimulus; it approached zero as learning progressed to asymptote. The U task had an average stimulus uncertainty of 0.84 bit per stimulus throughout the sessions. The R task, learned completely during the reception day, had a stimulus uncertainty of zero on all testing days. The *S*s were told that the S task contained a complex, but learnable structure, that the U task contained no identifiable pattern, and that the R task was completely predictable.

Results

Effect of signal uncertainty.—Figure 1, based on Group 1, shows that the three tasks, S, U, and R are sensitive to sleep loss, and that the error scores (missed signals) return to base-line levels after one night of recovery sleep. Sleep-loss decrement is considerably greater for the U task than for the other two.

A rank-order test (Lubin, 1961) performed on the base-line means of the 3-day and 5-day practice groups showed no significant decline in the percentage of errors of omission over the base-line days for either Group 1 or Group 2. Therefore, it was assumed that *S*'s performance was approximately stationary. An average base-line score (B) was computed for each *S*. The effect of the first night of sleep loss was estimated from $d_1 = D_1 - B$, the percentage of errors of omission after the first night of

[4] The apparatus is a modification of one designed by C. S. Gersoni.

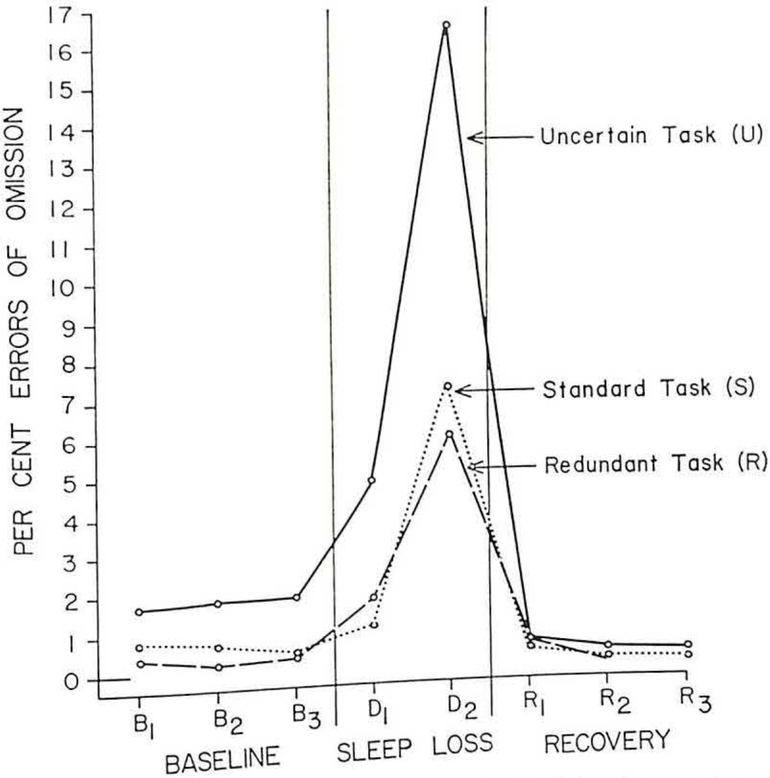


FIG. 1. Effect of sleep loss on visual vigilance with three levels of signal uncertainty (Individual Testing). (For the 5-day practice group, the last 3 base-line days were used.)

TABLE 1

SLEEP LOSS AND TYPE OF TASK

Task	d_1 (31 Hr.)		d_2 (55 Hr.)	
	M	Variance	*M*	Variance
Group 1 (Individual Test)				
Standard (S)	.81	7.94	5.99***	75.88
Uncertain (U)	2.95**	31.46	14.55***	338.42
Redundant (R)	1.25*	12.09	5.44***	71.94
Group 2 (Group Test)				
Standard (S)	.70	5.54	3.12**	58.56
Uncertain (U)	.84**	2.68	3.85***	31.10
Redundant (R)	.08	.62	.79*	4.17

Note.—$d_i = D_i - B$ where D_i is the percent errors of omission on the *i*th day of sleep loss, *B* is the average percent errors of omission during the base-line period.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

sleep loss minus the average base-line score. Similarly, the effect of 2 nights of sleep loss was estimated by $d_2 = D_2 - B$.

Table 1 shows average decrement scores and their variances on the three tasks for both Group 1 and Group 2. The *t* test[5] on these means showed that after 1 night of sleep loss, errors of omission on the U task increased significantly in both groups. After the second night of sleep loss, all three tasks showed a significant increase over base line in percentage errors of omission. Since the S, U, and R means were correlated, the Hotelling T^2 zero-mu test (Rao, 1952, p. 243) for the deviation of a vector of means from zero was performed within each

[5] All *t* tests in this paper are one-tailed.

group. It confirmed the univariate t tests.

These results implied that the U task was more sensitive to sleep loss than the other two. The Hotelling T^2 symmetry test (Rao, 1952, p. 241) for equality of correlated means showed that the average decrement scores for the three tests did not differ significantly after 1 night of sleep loss. However, after 2 nights of sleep loss, the impairment on the U task was significantly greater ($p < .01$) than on the other two tasks. This differential effect was found in both Group 1 and Group 2. In each group, the Hotelling linear function for the optimal measurement of sleep-loss impairment gave positive weight to U and negative weights to the S and R tasks for both the d_1 and d_2 vectors. The symmetry test thus confirmed the greater sensitivity of the U task to sleep loss.

To determine if there was an interaction between the *amount* of sleep loss and signal uncertainty, d_1 was subtracted from d_2 for each S on each task. In both Group 1 and Group 2 a Hotelling T^2 symmetry test on these vectors of difference means showed that the increase in errors of omission (from D_1 to D_2) was significantly greater ($p < .05$) for the U task than for the other two tasks.

Effect of task duration.—In Fig. 2, percentage errors of omission during each minute of the S task has been plotted against task duration for each sleep-loss day and for the base-line phase. Task-duration scores were available only for Group 1. Results for the recovery phase have not been plotted since they were almost exactly the same as the base-line scores. The U and R tasks (not shown) gave approximately the same results as the S task.

After 1 night of sleep loss, there was little change on the S task for the first 7 min. During the last 3 min., errors of omission increased rapidly over base-line levels. After 2 nights of sleep loss, omissions increased rapidly after 2 or 3 min. of work.

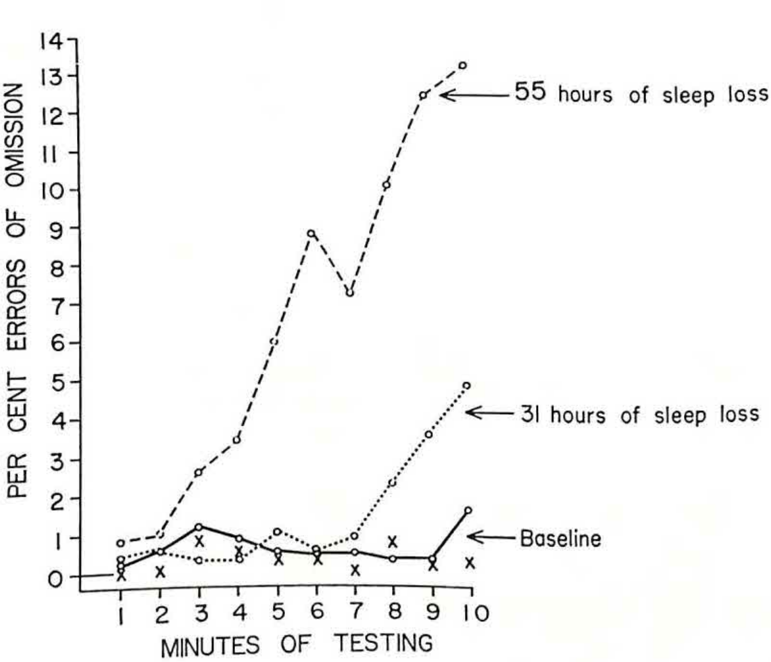


FIG. 2. Effect of duration and sleep loss on the Standard (S) task.

TABLE 2

TASK DURATION, TYPE OF TASK, AND SLEEP LOSS

	d_1 (31 Hr.)			d_2 (55 Hr.)		
	S	U	R	S	U	R
First 3 min.						
M	−.23	1.16	.93	.90	10.36	4.31
s^2	2.67	10.39	10.64	7.11	246.41	63.91
t	—	1.76*	1.39	1.65	3.23***	2.64**
Last 3 min.						
M	2.92	5.15	1.07	11.26	15.22	7.06
s^2	45.84	96.62	19.21	183.47	354.35	114.38
t	2.11*	2.57**	1.20	4.07***	3.96***	3.23***
Last minus first 3 min.						
M	3.15	3.99	.14	10.36	4.86	2.75
s^2	39.55	61.90	25.29	159.56	144.97	67.87
t	2.46*	2.49*	<1	4.02***	1.98*	1.86*

Note.—Entries in the table are percent errors of omission based on $d_i = D_i - B$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 2 shows that during the first 3 min. there was significant decrement on the U task after both the first and second sleep-loss nights. During the last 3 min., 1 night of sleep loss was sufficient to produce significant decrement on both the U and S tasks. After 2 nights of sleep loss all three tasks showed highly significant decrement.

The task-duration effect was estimated for each *S* by comparing the decrement score in the final 3 min. with the decrement score in the first 3 min. of each task. The bottom section of Table 2 shows that on all tasks, the percent errors of omission increased from first to last 3 min., and that, in general, the task-duration effect increased as sleep loss increased. The difference in decrement scores between first and last 3 min. was significant after 1 night of sleep loss for Tasks U and S. Task R, however, showed a significant duration effect only after 2 nights of sleep loss. The increase in the task-duration effect from 31 hr. to 55 hr. of sleep loss was significant only for Task S.

Effect of group testing.—Table 1 indicates that sleep-loss impairment was considerably less for Group 2 (group testing) than for Group 1. The *t* tests showed that the effect of the social situation was significant for the U and R tasks after the first night of sleep loss and for all three tasks after 2 nights without sleep. A linear discriminant function analysis confirmed this finding.

Effect of oral temperature.—Oral temperatures were taken on each *S* tested in group performance just before the testing session. The average base-line temperature was 97.99° F., decreasing to 97.96° F. after 1 night of sleep loss, and to 97.81° F. after 2 nights of sleep loss. These decreases

from base line were not significantly different from zero. The Pearson product-moment correlations of oral temperature changes (deviations from base-line average oral temperature) with the decrement scores after the first night of sleep loss were 0.09 for Task S, −0.19 for Task U, and −0.37 for Task R. A negative correlation indicated that errors of omission increased as oral temperature decreased. However, after 2 nights of sleep loss the correlations were all positive, and ranged from 0.08 to 0.29. Therefore, there was no consistent relation between oral temperature and errors of omission during sleep loss.

Discussion

The present findings confirm previous results obtained in this laboratory: increasing task uncertainty lowers the efficiency of the sleep-deprived *S*. Are these results, then, inconsistent with the earlier findings that complex tasks are not sensitive to sleep loss?

Wilkinson (1964) abandoned his preliminary guess that increased predictability potentiates performance decrement during sleep loss. Instead he suggested that motivation may be a key factor in the different kinds of results obtained. He assumed that tasks vulnerable to sleep loss largely lack incentive. Interesting, challenging tasks, on the other hand, resist the effects of sleep loss. Ordinarily, interesting tasks have high stimulus-response uncertainty, but uncertainty and interest have antagonistic effects during sleep loss. Unfortunately, no experiment has been done in which interest and uncertainty were varied independently.

Lapse hypothesis.—The impairment of performance during states of sleep deprivation has been attributed to the occurrence of lapses (Warren & Clark, 1937). Williams, Lubin, and Goodnow (1959) and Williams, Granda, Jones, Lubin, and Armington (1962) showed that lapses (defined on the basis of the EEG as brief periods of light sleep) were associated with errors of omission in work-paced vigilance tasks and that lapses increased in frequency, duration, and depth as sleep loss increased.

Since greater signal uncertainty causes longer S-R intervals (Hyman, 1953), the relatively high impairment in performance on the U task may have resulted from a greater coincidence of lapses with critical periods in the protracted S-R interval. Assuming that the probability of a lapse and the probability of a signal are independent, frequency of missed signals should be a multiplicative function of the two; therefore, one would expect, as found here, an ordinal interaction between sleep loss and signal uncertainty. Further analysis of the data of Williams et al. (1962) revealed also that during sleep deprivation the frequency of EEG lapses increased as the task duration increased. Thus, the increase in errors of omission with increasing task duration could be the result of an increase in the rate of lapses. Finally a beneficial effect of interest on task performance may result from the occurrence of a lower lapse rate during the performance of interesting tasks.

Social stimulation has been shown previously to offset the effects of sleep deprivation (Ax, Fordyce, Loovas, Meredith, Pirojnikoff, Shmavonion, & Wendalh, 1953; Laties, 1961). The results of the present experiment, where *S*s working in a group showed less decrement than *S*s tested individually, confirm these findings.

There was no relation between change in oral temperature and errors of omission. Thus, the present data provide an exception to Kleitman's generalization that performance follows body temperature (Kleitman, 1933).

References

Adams, J. A., Stetson, H. H., & Humes, J. M. Monitoring of complex visual displays: II. Effects of visual load and response complexity on human vigilance. *Hum. Factors*, 1961, **3**, 213–221.

Ax, A. F., Fordyce, W., Loovas, I., Meredith, W., Pirojnikoff, L., Shmavonion,

B., & Wendalh, R. Quantitative effects of sleep deprivation. *USA Qm. Res. Develpm. res. final Rep.*, 1953, No. DA 44-109-Qm-1130.

Broadbent, D. E. The twenty dials test under quiet conditions. *Med. Res. Council, Appl. Psychol. Res. Unit Rep., Great Britain*, 1950, No. AP4 130/50.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Katz, S. E., & Landis, C. Psychologic and physiologic phenomena during a prolonged vigil. *Arch. Neurol. Psychiat.*, 1935, **34**, 307–317.

Kleitman, N. Studies on the physiology of sleep: VII. Diurnal variation in performance. *Amer. J. Physiol.*, 1933, **104**, 449–456.

Laties, V. Modification of affect, social behavior, and performance by sleep deprivation and drugs. *J. psychiat. Res.*, 1961, **1**, 12–25.

Lubin, A. L'utilisation des corrélations par rang pour éprouver une tendance dans un ensemble de moyennes. *Bull. Cent. Etud. Rech. Psychotech.*, 1961, **10**, 433–444.

McGrath, J. J., Harabedia, A., & Buckner, D. N. Review and critique of the literature on vigilance performance. Technical Report No. 1, December 1959, Human Factors Research Incorporated, Contract Nonr 2649(00), Office of Naval Research.

Montague, E. K., Williams, H. L., Lubin, A., & Gieseking, C. F. The use of Army tests for the assessment of intellectual deficit. *U. S. Armed Forces med. J.*, 1957, **8**, 883–892.

Rao, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

Robinson, E. S., & Robinson, F. Effects of sleep loss. II. *J. exp. Psychol.*, 1922, **5**, 93–100.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., Jr., & Beck, L. H. A continuous performance test of brain damage. *J. consult. Psychol.*, 1956, **20**, 343–350.

Warren, N., & Clark, B. Blocking in mental and motor tasks during a 65-hour vigil. *J. exp. Psychol.*, 1937, **21**, 97–105.

Wilkinson, R. T. The effects of sleep loss on performance. *Med. Res. Council, Appl. Psychol. Res. Unit Rep., Great Britain*, 1958, No. 323.

Wilkinson, R. T. Effects of up to 60 hours of sleep deprivation on different types of work. *Ergonomics*, 1964, **7**, 175–186.

Williams, H. L., Granda, A. M., Jones, R. C., Lubin, A., & Armington, J. C. EEG frequency and finger pulse volume as predictors of reaction time during sleep loss. *EEG clin. Neurophysiol.*, 1962, **14**, 64–70.

Williams, H. L., Lubin, A., & Goodnow, J. J. Impaired performance with acute sleep loss. *Psychol. Monogr.*, 1959, **73**(14, Whole No. 484).

(Received February 17, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 408–412

S-R COMPATIBILITY AND INFORMATION REDUCTION[1]

PAUL M. FITTS AND IRVING BIEDERMAN

University of Michigan

S-R compatibility effects were examined in 4 information-processing tasks (1-bit information conserving, 2-bit conserving, 2 to 1 bit filtering, and 2 to 1 bit condensing) in combination with 2 sets of responses (2 or 4 fingers of 1 hand only vs. 1 or 2 fingers of both hands). 8 different groups of 10 Ss each were used, 1 under each condition, and tested for 2 sessions. 1-bit conserving and 2 to 1 bit filtering were accomplished about equally well, under both response codes. The other 4 tasks involved significantly more time and errors. When a compatible (2-hand) response code was used, 2-bit information conserving was more efficient than 2 to 1 bit information condensing, notwithstanding the fact that the former involved twice as many alternative responses; these relations were reversed when a less compatible (1-hand) response code was used. These results indicate the importance of response coding in interpreting studies of different information-handling processes.

Morin, Forrin, and Archer (1961) recently investigated the effects of different "perceptual demands" imposed by irrelevant stimulus information in tasks where *S*s are required to use fewer responses than there are stimuli, and concluded that performance is essentially independent of variations in perceptual demands, being primarily a function of transmitted information. Posner (1962, 1964) has subsequently proposed a taxonomy for such tasks. He proposes the term information conservation for tasks in which the number of stimulus and response categories correspond, and the term information reduction for tasks in which there are fewer response categories than stimulus categories. Information reduction may be accomplished by *filtering* (or gating), where one or more stimulus dimensions are irrelevant and may be ignored, or else by *condensing* information, where all stimulus dimensions are relevant and must be considered. Condensing information would appear to impose greater perceptual demands than filtering. Thus Fitts (1959) has predicted that filtering should, in most cases, be accomplished more efficiently than information conservation, given the same stimulus source, whereas Posner (1962), examining information-condensing performance in a variety of tasks, has found that the latter form of information reduction requires more time than does information conservation, given the same sequence of stimuli. Morin, Forrin, and Archer's (1961) results are an exception to these generalizations, since condensing as well as filtering information (Cond. IV and V of Fig. 1) was found to be faster than information conservation (Cond. III).

In examining the Morin, Forrin, and Archer (1961) experiment, it appeared to the present authors that their four-choice, information-conservation task might be low in S-R compatibility (see Fitts & Seeger, 1953), and that compatibility effects might thus account for the fact that this was

[1] This research was supported by the Advanced Research Projects Agency, Department of Defense, and was monitored by the Air Force Office of Scientific Research, under Contract No. AF 49(638)-1235 with the Human Performance Center, Department of Psychology.

GROUP	INFORMATION PROCESSING TASK	STIMULUS-RESPONSE CODES			
		Stimuli			
		●●	●	■	■■
		Responses*			
II	Conservation(1—1 Bit)		1-R	2-R	
II-B	Conservation(1—1 Bit)		1-L	1-R	
III	Conservation(2–2 Bits)	3-R	1-R	2-R	4-R
III-B	Conservation(2–2 Bits)	2-L	1-L	1-R	2-R
IV	Filtering (2—1 Bit)	1-R	1-R	2-R	2-R
IV-B	Filtering (2—1 Bit)	1-L	1-L	1-R	1-R
V	Condensing (2—1Bit)	2-R	1-R	2-R	1-R
V-B	Condensing (2—1Bit)	1-R	1-L	1-R	1-L

*1 = First or forefinger, 2 = Second or middle finger, etc.
L = Left hand, R = Right hand.

FIG. 1. Experimental conditions for the eight groups.

the most difficult of the four tasks which they studied. Accordingly it was decided to try to find a more compatible set of four-finger responses for use with their stimuli (one square, one circle, two squares, or two circles).

Population stereotypes were determined by asking *S*s to indicate the four fingers they would prefer to use with the four stimuli and how they would form S-R pairs. Out of approximately 100 *S*s so queried, nearly all selected the middle and forefingers of the right and left hands, assigning stimulus *form* to hands, and stimulus *number* to fingers as shown in Cond. III-B, Fig. 1. This "two-dimensional" response set was therefore selected as likely to be more compatible than the one-hand response set (Cond. III) used in the previous study. All four information-handling tasks were studied, using both one-hand and two-hand response sets, but it was predicted that S-R compatibility effects should be prominent only for the task requiring the use of four responses (Cond. III and III-B), since result from recent studies (see Broadben 1963; Fitts, 1964; Griew, 1958) in dicate that compatibility effects ar negligible for two-alternative choice but increase rapidly as the number c alternatives is increased, i.e., con patibility effects show up in the *slop* of the function relating reaction tim (RT) to information transmitted, nc in the intercept.

METHOD

Stimulus forms were drawn in black ink c white cards, and exposed by the opening of double-bladed shutter. Exposure was te minated by *S*'s response, whether correct c in error (instead of by correct response onl as in the previous study), and RTs we measured for the initial response, wheth correct or in error (instead of for the correctic when an error was made), in order to confor to conventions for computing informatio transmission rates. Otherwise the meth employed by Morin, Forrin, and Arch (1961) was used.

Ten *S*s were assigned randomly to each eight groups, and each group assigned to o of the conditions of Fig. 1. All *S*s were test first for one block of trials under simple R

conditions (Cond. I), as in the previous study, and were then given seven blocks of choice RT trials, 16 trials per block, on one session, followed by eight blocks of 16 trials during another session 48 hr. later.

Results

Reaction time.—Median RTs were computed for each *S* for each block of trials, and group means determined. The eight tasks clearly fall into two distinct clusters (see Fig. 2). Throughout both sessions Cond. III and V required longer RTs than did Cond. II and IV. Since there was an overlap of only 3 *S*s in the tails of the two RT distributions out of a total of 80 *S*s, the significance of this gross difference is clear. Data for the two sets of conditions and for each session were therefore analyzed separately.

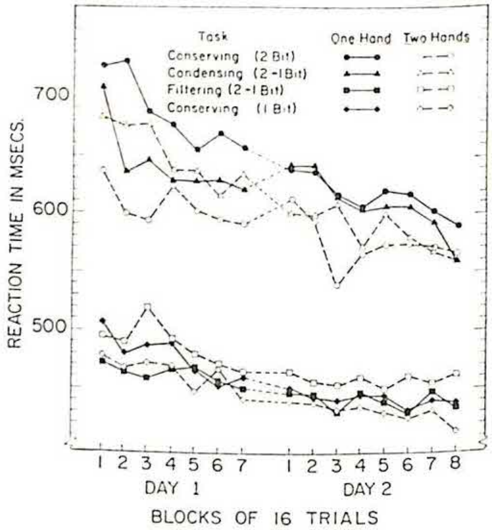


FIG. 2. Group means for the eight experimental conditions, based on individual medians for each block of 16 trials.

As stated previously, comparisons of performance among the four most difficult tasks is most relevant for assessing S-R compatibility effects. Inspection of Fig. 2 reveals that the predicted effects are in fact present. The shift to a more compatible two-hand response set reduced the mean RT for the 2-bit information-conservation task by about 100 msec. on initial trials and reversed the superiority of the 2 to 1 bit information-condensing task relative to the 2-bit information-conserving task. Results of an analysis of variance for these data (see Table 1) indicate that

TABLE 1

Analyses of Variance for the Difficult Tasks for Both Sessions

Source	Day 1			Day 2		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *S*s	39			39		
Info. proc. task (I)	1	.7	*ns*	1	3.9	*ns*
Response codes (C)	1	885.7	4.51*	1	849.9	4.80*
I × C	1	1395.2	7.11*	1	140.8	*ns*
*S*s within groups (*S*s)	36	196.2		36	177.1	
Within *S*s	240			280		
Trials (T)	6	203.7	6.23**	7	127.2	6.91**
I × T	6	20.1	*ns*	7	22.1	*ns*
C × T	6	20.2	*ns*	7	9.6	*ns*
I × C × T	6	47.6	*ns*	7	17.9	*ns*
T × *S*s	216	32.7		252	18.4	

* $p < .05$.
** $p < .01$.

TABLE 2

ERROR AND INFORMATION TRANSMISSION DATA FOR THE EIGHT GROUPS OVER BOTH SESSIONS

Group	Task	Response Code	Percent Errors	H_t per Response (Bits)	H_t per Second (Bits)
II	1-bit conserv.	1 hand	3.62	0.81	1.76
II-B	1-bit conserv.	2 hand	4.33	0.76	1.70
III	2-bit conserv.	1 hand	9.04	1.56	2.40
III-B	2-bit conserv.	2 hand	7.79	1.63	2.76
IV	2 to 1 bit filter.	1 hand	4.79	0.77	1.70
IV-B	2 to 1 bit filter.	2 hand	5.00	0.77	1.63
V	2 to 1 bit condens.	1 hand	7.96	0.69	1.10
V-B	2 to 1 bit condens.	2 hand	11.75	0.51	0.83

learning (trials) was statistically significant within each day, that response coding effects (two vs. one hand) were significant within each day, and that the Task × Response Code interaction was significant on the first but not on the second day. The last result indicates that the magnitude of compatibility effects decreased with training.

An analysis of variance for the four easier tasks indicated that neither the use of one vs. two hands, nor filtering vs. conserving of information had a significant main effect. However, learning was again highly significant, as was the interaction of Trials × Experimental Conditions.

Errors.—Errors for the eight conditions (see Table 2) were less frequent than in the earlier study by Morin, Forrin, and Archer (1961). Mean errors correlated .87 with mean RTs across the eight tasks, and the four tasks with longest RTs all had the highest error rates. Analyses of variance for the error scores for the four most difficult conditions and for the four easier conditions indicated that within these sets of data differences among tasks were not statistically reliable. However, within the four most difficult tasks error rates followed a similar interaction pattern as did the RT scores; in the 2 to 1 bit information-condensing task more errors were made with two hands than with one, with the reverse being true for the 2-bit information-conserving task. The error data thus lend support to the findings based on the RT data.

Information-transmission rates.—Information-transmission rates (see Table 2) for the 2 to 1 bit information-condensing task were by far the lowest of any of the tasks studied, the average value of information transmission rate (H_t/sec) being barely half that for 2 to 1 bit information filtering or for 1-bit conservation.

DISCUSSION

For three of the four conditions that constitute a replication of Morin, Forrin, and Archer's (1961) experiment, the present results agree fairly closely with the original data. However, this is not true for Cond. V. In the earlier study the information-condensing task, Cond. V, gave relatively faster RTs than those obtained in our replication, and also gave much faster RTs than those obtained for the 10 *S*s tested under Cond. V-B of our experiment. One explanation is that of an unusually discrepant sample of *S*s in the earlier study; this explanation is given credence by the fact that Morin, Forrin, and Archer (1961) reported that this group contained ". . . two highly atypical *S*s who responded very rapidly and transmitted information with near-perfect accuracy [p. 94]." The present findings indicate a large superiority of

the 2 to 1 bit filtering task over both the 2 to 1 bit condensing and the 2 to 2 bit information-conserving tasks. However, considering both sessions, the rank order of the four replicated tasks is the same as in the original study. Only when the two-hand response set is used is the rank of the two most difficult tasks reversed, this effect being statistically significant on the first experimental session.

The present findings therefore support the conjecture that the results obtained by Morin, Forrin, and Archer (1961) were due in part to the choice of a response set which did not permit the formation of a compatible S-R ensemble. In view of these results, it therefore does not appear wise to conclude, as they do, that "apparently some minimum amount of time is needed to learn new or unusual codes but once they are learned, coding may become a relatively unimportant determinant of the rate of information processing [p. 96]," or that "the rate at which information can be processed can adequately be accounted for in terms of formal information measures without reference to the perceptual demands of the task [p. 95]." Instead it appears that performance does depend in part on the "perceptual demands" of different information-processing tasks, as Morin, Forrin, and Archer (1961) originally surmised.

The present results, considered in conjunction with earlier data (Alluisi & Muller, 1958; Brainard, Irby, Fitts, & Alluisi, 1962; Broadbent, 1963; Broadbent & Gregory, 1962; Deininger & Fitts, 1955; Fitts & Seeger, 1953; Leonard, 1961), suggest that efficiency in information reduction, as well as in information-conserving tasks, is determined in part by stimulus-response interactions, or S-R compatibility effects.

The examination of human performance in tasks that require selective information reduction, such as filtering and condensing information, would appear to be an important bridge to an understanding of more complex perceptual and cognitive processes. As an illustration, it appears that the number of dimensions of the stimulus and response is an important determinant of information-processing performance, as has frequently been shown to be the case in concept-identification and classification-learning studies.

REFERENCES

ALLUISI, E. A., & MULLER, P. F. Verbal and motor responses to seven symbolic codes: A study in S-R compatibility. *J. exp. Psychol.*, 1958, **55**, 247–254.

BRAINARD, R. W., IRBY, T. S., FITTS, P. M., & ALLUISI, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

BROADBENT, D. E. S-R compatibility and the processing of information. In, *Proceedings of the XVII international congress of psychology.* Amsterdam, Netherlands: North-Holland, 1963.

BROADBENT, D. E., & GREGORY, M. Donders' B- and C-reactions and S-R compatibility. *J. exp. Psychol.*, 1962, **63**, 575–578.

DEININGER, R. L., & FITTS, P. M. Stimulus-response compatibility, information theory, and perceptual-motor performance. In H. Quastler (Ed.), *Information theory in psychology: Problems and methods.* Glencoe, Illinois: Free Press, 1955.

FITTS, P. M. Human information handling in speeded tasks. Report No. RC-109, June 24, 1959, IBM Research Center, Yorktown Heights, New York.

FITTS, P. M. Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

FITTS, P. M., & SEEGER, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

GRIEW, S. Information gain in tasks involving different stimulus-response relationships. *Nature*, 1958, **192**, 1819.

LEONARD, J. A. Choice reaction time experiments and information theory. In C. Cherry (Ed.), *Information theory.* London: Butterworth, 1961.

MORIN, R. E., FORRIN, B., & ARCHER, W. Information processing behavior: The role of irrelevant stimulus information. *J. exp. Psychol.*, 1961, **61**, 89–96.

POSNER, M. I. *An informational approach to thinking.* (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1962, No. 63–433.

POSNER, M. I. Information reduction in the analysis of sequential tasks. *Psychol. Rev.*, 1964, **71**, 491–504.

(Received February 18, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 413–418

DISCRIMINATION PERFORMANCE AS AFFECTED BY PROBLEM DIFFICULTY AND SHOCK FOR EITHER THE CORRECT OR INCORRECT RESPONSE [1]

HARRY FOWLER AND GEORGE J. WISCHNER

University of Pittsburgh

To assess one interpretation of the paradoxical facilitating effect of shock for the correct response, 12 groups of 8 hungry rats each received visual-discrimination training for food under no-shock, shock-right, and shock-wrong conditions and 4 levels of problem difficulty. Difficulty was manipulated by varying the relative brightnesses of the discriminanda. Performance was progressively retarded with increasing levels of difficulty. However, with the more difficult discriminations, performance relative to that of the no-shock Ss was facilitated by the shock-right condition and more so by the shock-wrong condition. It is suggested that shock may have both an avoidance and cue function, the latter presumably providing the basis for the finding, obtained for the 1st time in our laboratory, that shock for the correct response facilitates performance.

The present study was designed to assess the effect of shock for either the correct or incorrect response on performance in visual-discrimination tasks varying in level of difficulty. The rational for investigating the difficulty-of-discrimination variable stems from one interpretation of the paradoxical facilitating effect of shock for the correct response examined by Freeburne and Taylor (1952). These investigators, suggesting that shock, associated only with the correct response, may serve as a differential cue and hence as a secondary reinforcer, trained animals on a task involving shock for both correct and incorrect responses, a condition precluding any cue effect of the shock. The finding that this shock-both group performed significantly better than a no-shock group led Freeburne and Taylor to reject the cue-function interpretation of the facilitating effect of shock for the correct response and to interpret their data as supporting a general sensitizing function of shock as originally proposed by Muenzinger (1934). Prince (1956) questioned Freeburne and Taylor's (1952) results on certain methodological grounds, however, and failed in his experiment to obtain any difference between shock-both and no-shock groups.

In a very recent study of the effect of various combinations of food and shock for either the correct or incorrect response, Wischner, Hall, and Fowler (1964), also obtained no difference between shock-both and no-shock groups. More important was the finding that a shock-right group, a control absent in both the Prince (1956) and Freeburne and Taylor (1952) studies, performed significantly poorer than the no-shock group. This failure to obtain a facilitating effect of shock for the correct response is in accord with our previous findings, indicating that shock for the correct response serves

[1] This study was supported in part by Grant No. G-14312 from the National Science Foundation and Grant No. MH-08482 from the National Institute of Mental Health, United States Public Health Service. Grateful acknowledgment is made by the authors to D. Bernstein and N. Ketterle for their assistance in the conduct of this experiment.

only to retard performance (Wischner, 1947), the extent of this retardation being greater with increasing intensities of shock (Wischner, Fowler, & Kushnick, 1963).

The present investigation examines further the possibility that the facilitating effect of shock for the correct response reported by Muenzinger and his colleagues (Muenzinger, 1934; Muenzinger, Bernstone, & Richards, 1938) may relate to a possible cue function of the shock. It is proposed that such an effect might be obtained with relatively difficult discriminations in which the alternatives, including the discriminative stimuli, are highly similar. Under these conditions, shock for either the correct or incorrect response may serve to heighten the discriminability of the alternatives, thereby delimiting any generalization of secondary reinforcement from the correct (food-reinforced) goal arm to the incorrect arm, and thus reducing errors.

Method

Subjects.—The *S*s for the experiment were 96 naive, male albino rats of the Sprague-Dawley strain, about 100 days old at the start of the experiment. The *S*s were caged individually in the experimental room under controlled temperature and an artificially illuminated day-night cycle.

Apparatus.—The apparatus was an enclosed T maze with a stem 26 in. long, and each arm, as measured from the center of the choice point, 36 in. long. The height of each section of the maze was uniformly 4½ in. and the width 3¾ in. The initial 8-in. section of the stem was separated from the remaining portion by a guillotine door and served as the start compartment. Guillotine doors were also positioned near the end of each arm, to provide 12-in. goal compartments, and at the entrance to each arm, 3 in. from the center of the choice point, to prevent correction.

Each maze section was constructed of two L-shaped strips of galvanized sheet metal, an exterior supporting shell of wood, painted flat black, and a transparent Plexiglas top. One L-shaped strip of sheet metal served as one wall and half of the floor of the maze section; the other L served as the other half of the floor and the other wall. Together, the two L-shaped strips provided two, 1½-in. floor surfaces which were separated by a ¾-in. gap.

Each goal compartment had a clear Plexiglas food well (¾ in. in diameter × ½ in. deep) and an end plate or wall made of frosted Plexiglas. The frosted, goal-box end plates were visible from the choice point and served as the discriminative stimuli when each was differentially illuminated by a frosted 10-w. bulb located directly behind the end plate in a wooden enclosure painted a flat black.

To effect different levels of problem difficulty, the brightnesses of the two goal-box end plates were varied by operating one of the goal-box bulbs at 120 v. and the other at either 96, 87, 78, or 63 v. These four sets of voltages for the dim and brightly illuminated goal-box end plates provided footcandle ratios of .50, .37, .25, and .12, respectively. Hereafter, reference will be made to these four sets of differential goal-box brightnesses as the Difficult (D), Medium Difficult (MD), Medium (M), and Medium Easy (ME) problem conditions, respectively. Additional illumination, exterior to the maze, was provided by two 10-w. bulbs (120-v. source), one located near the end of each arm on the stem side of the maze.

A matched-impedance shock system consisting of a 60-cps ac source and a series resistance of .3 megohms was used to deliver shock to *S* in either maze arm. In this shock system, the two L-shaped strips of sheet metal forming the sides and floor of the maze were connected across the output of the transformer; thus, *S* received shock when it made contact with both halves of the sheet-metal floor. Because of the narrowness of the maze alley, *S* could not avoid shock by running along only one half of the floor surface.

Shock was delivered to *S* when it interrupted in either arm, an infrared photoelectric beam crossing the arm at a point midway between its two guillotine doors. A manual priming feature of the shock circuit prevented *S* from receiving more than one shock in any one trial, even though it was possible for *S* to interrupt a photoelectric beam several times during a trial. When the shock was delivered, its duration was held constant at .2 sec., as metered through a Hunter timing relay, and its intensity was set at 60 v., as measured across the output of the transformer.

Procedure.—The experimental procedure included both pretraining and training phases. One week prior to the pretraining phase, *S*s were started, and then maintained for the

duration of the experiment, on a daily diet of 12 gm. Purina lab checkers, with water available ad libitum. During this week each *S* was handled daily for about 3–4 min.

For both the pretraining and training trials, *S*s were kept in individual compartments of a hardward cloth detention cage. Prior to their daily run in the maze, which took place during the dark phase of the artificial day-night cycle, all *S*s received approximately 5 min. adaptation to the conditions of illumination of the room as provided by the two exterior 10-w. apparatus bulbs.

Pretraining was administered to habituate *S* to the apparatus and to reduce possible position and brightness preferences. Since the latter would presumably relate to the different brightnesses of the discriminative visual stimuli, the 96 *S*s were randomly assigned to four equally numbered groups, each of which was run, during both pretraining and training, under one of the four different problem conditions.

For pretraining, each *S* received a total of 16 forced, reinforced trials, administered 4 per day and randomly distributed with the restriction that half of the trials were to the left and half to the right, with half of each of these sets being to a bright goal and half to a dim goal. Forcing was accomplished by lowering one of the guillotine doors at the choice point. The reinforcement provided on each trial consisted of P. J. Noyes Formula A rat pellets (4 mm., 45 mg.). On the first of the 4 pretraining days, the food pellets were liberally spread throughout the goal box; on subsequent days, the number and locus of pellets were systematically reduced until, on the last day, two pellets were given only in the goal-box food cup. With each forced trial, *S* was permitted to run to the available goal where it was detained until all the pellets were consumed or until 5 min. had elapsed. Then, *S* was removed from the maze and replaced in the carrying cage to await its next trial. On the average, the interval between successive trials within a day was about 15 min., the time required to run a block of *S*s.

Training began on the day following termination of pretraining and consisted entirely of free-choice trials. For the training phase, the 24 *S*s of each of the four problem-difficulty conditions were randomly assigned to three equally numbered shock subgroups: shock-right (SR), shock-wrong (SW), and no-shock (NS). Shock-right *S*s received both shock and food reinforcement for a correct response but neither for an incorrect response; shock-wrong *S*s received food reinforcement for a correct response and shock for an incorrect response; the control, no-shock *S*s received only food for a correct response.

All *S*s received four trials per day for the first 6 days of discrimination training and eight trials per day thereafter. During these trials, food reinforcement, consisting of two pellets, could be obtained only in the bright goal, the right-left position of which was varied throughout training in accordance with a predetermined random schedule. On any training trial, detention time in either the correct or incorrect goal was approximately 10 sec., and the interval between successive trials within a day about 15 min., again the time required to run a block of *S*s. Discrimination training was continued until each *S* met a criterion of 15 correct responses out of a total of 16 free choices with the last 8 being correct, or until a total of 400 training trials had been administered.

Results and Discussion

Group mean errors in blocks of 20 trials are presented in Fig. 1. The separate panels of this figure provide the mean error scores for the SR, SW, and NS groups within each of the problem-difficulty levels. As shown, mean errors decreased over the course of training for all groups except the NS group of the difficult problem. Overall, the rate of error reduction tended to be progressively slower, the more difficult the discrimination task. More important is the fact that, within each difficulty level, both the SR and SW *S*s showed more rapid elimination of errors than did the control, NS *S*s.

Because of the generally orderly progression of the performance curves for the three shock groups within each difficulty level, only group mean errors for the entire 400 trials were subjected to statistical analysis. These data are presented for the three shock groups in Fig. 2 as a function of task difficulty. Included in this figure for the "E" or easy problem are data from a previous study (Wischner et al., 1963). The significance of these additional data, presented for comparison purposes, will be considered below.

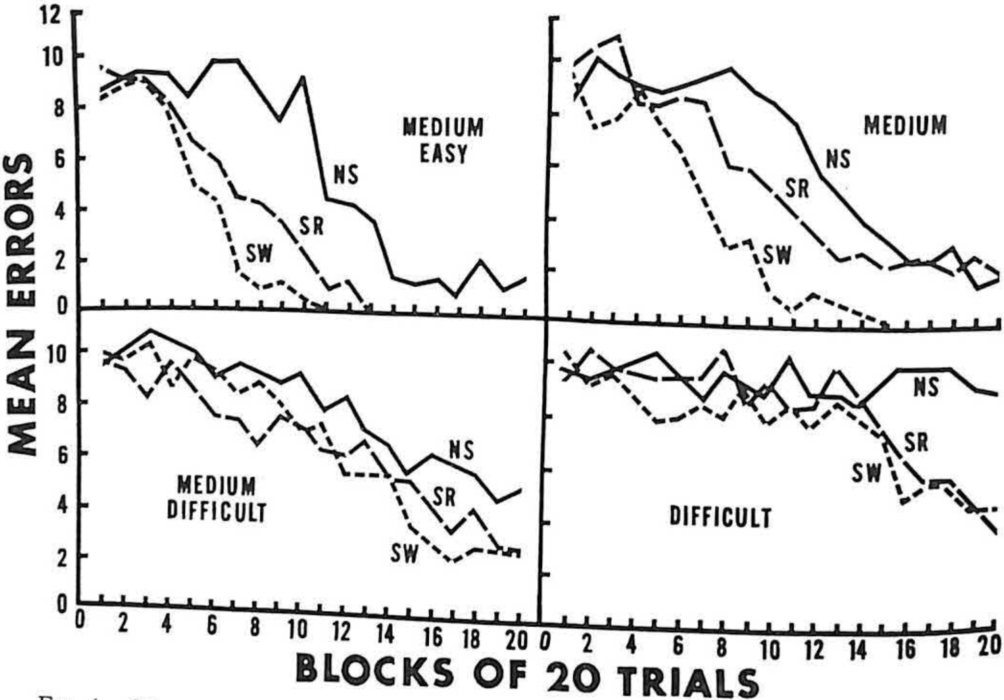


FIG. 1. Mean errors, per 20 trial blocks, for the shock-right (SR), shock-wrong (SW), and no-shock (NS) groups within each condition of problem difficulty.

The results of a 3×4 factorial analysis of variance of the data presented in Fig. 2 for the present study showed that group mean total errors were positively related to level of task difficulty; both the differences among the several difficulty levels and the linear trend in the means for these levels were highly reliable; $F\ (3, 84) = 37.2$, $p < .001$ and $F\ (1, 84) = 111.4$, $p < .001$, respectively. Residual trends after extraction of the linear component were all nonsignificant, $F < 1$. The overall differences among the three shock conditions were also highly reliable, $F\ (2, 84) = 16.6$, $p < .001$ and, as indicated by a nonsignificant Shock $\times$ Difficulty Level interaction, $F < 1$, related generally to the range of difficulty levels studied. In addition, separate comparisons, based on the overall error term, showed that the difference between any pair of shock groups was reliable: NS vs. SR, $F\ (1, 84) = 11.4, p < .001$; NS vs. SW, $F\ (1, 84) = 32.7, p < .001$; SR vs. SW, $F\ (1, 84) = 5.5, p < .025$.

In contrast to the data of the present investigation, those presented in Fig. 2 for the Wischner et al. (1963) study show that the performance of the SR group was retarded relative to that of the NS group. These additional data are for comparable groups of *Ss* run in the same apparatus under training conditions which, for all purposes, were identical to those employed in the present study, except for the fact that the problem employed was a light-dark discrimination rather than one of the present set of bright-dim discriminations. For this reason, these additional data have been presented in Fig. 2 as being representative of the effects of shock for either the correct or incorrect response on performance in an "easy" discrimination task. In connection with these data, it should be noted that other data from the Wischner et al. study, not included in

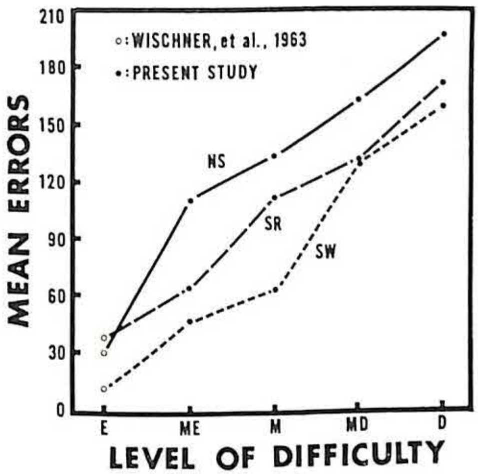


FIG. 2. Mean errors over the entire 400 training trials for the SR, SW, and NS groups as a function of level of problem difficulty. (The data presented for the "E" or easy problem level are for comparable groups of *Ss* run in a previous study [Wischner et al., 1963].)

Fig. 2, showed an even greater retardation of performance for SR animals when either a forced-choice training procedure or a higher intensity of shock was employed in their easy problem condition.

In total, the data presented in Fig. 2 suggest that, within the context of discrimination learning, shock may serve, at least, two functions. Through its aversive property, shock may produce avoidance of the stimuli with which it is associated, thereby increasing errors for SR *Ss* and reducing them for SW *Ss*. Through its cue property, however, shock for either the correct or incorrect response may increase the discriminability of the goal arms and thereby facilitate performance. This cue effect of shock should be particularly apparent when the discriminative stimuli are highly similar, as in the case of the difficult discriminations employed in the present study. With this arrangement, the secondary reinforcing effects of the cues present in the correct goal arm may generalize to those in the incorrect arm, to impede formation of the discrimination. But, when shock is consistently administered for one response, both the shock experience, itself, and any fear elicited by it may serve to make the two goal arms more distinctive, thereby reducing the generalization of secondary reinforcement from the correct to the incorrect goal arm and thus reducing errors. In contrast, when the discriminative stimuli are quite dissimilar, as in the simple or easy discrimination, the cue effect of shock should be relatively small, and in the case of the SR condition, possibly be offset or obscured by the shock's aversive or avoidance-producing property.

In view of the present results, and also the fact that Muenzinger and his colleagues consistently obtained facilitation with their SR conditions, certain aspects of Muenzinger's (1934) general procedure are particularly noteworthy. Muenzinger always employed a simple black-white discrimination, but his animals were required, following a correct choice, to run from the positive discriminative stimulus into a gray goal where food reinforcement was obtained. This training procedure would seem to enhance the "difficulty" of the discrimination, in so far as it promotes a temporal dissociation of the positive discriminative stimulus and the food reinforcement, and also affords a greater similarity between the gray goal-box cues associated with food and those present in the incorrect arm. Under these conditions, then, as in the present study, the SR condition could be expected to facilitate performance, since the shock experience, itself, and any fear elicited by it, would serve both to mediate the reinforcing effect of food and to minimize any generalized secondary reinforcing effects of the cues present in the incorrect arm.

Whether the above interpretation may fully account for Muenzinger's results is obviously a matter for additional study of the variables relevant to the problem-difficulty dimension. It should be apparent that, as related to the rate of acquisition of a discrimination, the dimension of problem difficulty may be manipulated by varying the types of factors suggested to be operative in Muenzinger's situation or by varying the similarity of the discriminanda, as in the present study. Aside from these con-

siderations, however, the findings of the present study seem to be of significance for general research in the area. Church (1963) has commented in his recent review of the varied effects of punishment that "an important contribution would be made by the identification of a parameter that would result in facilitation by punishment at some values and inhibition by punishment at other values [p. 388.]" The results of the present study indicate that the dimension of problem difficulty is such a parameter; its isolation now makes possible an assessment of the degree to which other variables may contribute either to the facilitating or disrupting effect of shock for the correct response.

REFERENCES

Church, R. M. The varied effects of punishment on behavior. *Psychol. Rev.*, 1963, **70**, 369–402.

Freeburne, C. M., & Taylor, J. E. Discrimination learning with shock for right and wrong responses in the same subject. *J. comp. physiol. Psychol.*, 1952, **45**, 264–268.

Muenzinger, K. F. Motivation in learning: I. Electric shock for correct responses in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 439–448.

Muenzinger, K. F., Bernstone, A. H., & Richards, L. Motivation in learning: VIII. Equal amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.

Prince, A. I., Jr. Effect of punishment on visual discrimination learning. *J. exp. Psychol.*, 1956, **52**, 381–385.

Wischner, G. J. The effect of punishment on discrimination learning in a non-correction situation. *J. exp. Psychol.*, 1947, **37**, 271–284.

Wischner, G. J., Fowler, H., & Kushnick, S. A. The effect of strength of punishment for "correct" or "incorrect" responses on visual discrimination performance. *J. exp. Psychol.*, 1963, **65**, 131–138.

Wischner, G. J., Hall, R. C., & Fowler, H. Discrimination learning under various combinations of food and shock for "correct" and "incorrect" responses. *J. exp. Psychol.*, 1964, **67**, 48–51.

(Received February 24, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 419–425

THE REINFORCEMENT RELATION AS A FUNCTION OF INSTRUMENTAL RESPONSE BASE RATE [1]

ROBERT W. SCHAEFFER [2]

University of Missouri

This experiment investigated the relationship between base rate (i.e., operant level, or precontingent rate) of the instrumental response and asymptotic reinforced responding. Running was used to reinforce an instrumental response of licking. Variations in the base rate of the instrumental licking response were produced by varying the contents of the drinking tube, while base rate of the reinforcing running response was held constant. The effects of variations in the base rate of the instrumental response on the reinforcement relation were examined for several FR reinforcement schedules and for extinction, both with respect to asymptotic reinforced licking, and with respect to the reinforced increments in contingent licking. The results indicated that asymptotic reinforced instrumental responding varied directly with base rate of the instrumental response, and that the reinforced increments in contingent instrumental responding were determined both by base rate of the instrumental response and the reinforcing response.

Although numerous studies have investigated the changes associated with acquisition and extinction of the instrumental response under a variety of reinforcement schedules, the possibility that the base rate (i.e., operant level, or precontingent rate) of the instrumental response might itself influence the outcome of the reinforcement relation has been largely ignored. Neglect of the instrumental response base line as a potential contributor to the reinforcement relation is undoubtedly related to the general theoretical conception that *some events are*, and *some events are not* reinforcers (i.e., some responses may be used only as instrumental responses, while other responses may be used only as reinforcing responses). Within this context, reinforcing responses have been treated as absolute, categorical, and transituational events (cf. Meehl, 1950; Miller & Dollard, 1941) which have well-established distributional properties and which, when made contingent upon a given instrumental response, actively alter the distributional properties of the instrumental response. Instrumental responses, on the other hand, have been treated solely as passive responses which, although they are *affected by* the reinforcing response, have no *effect upon* the reinforcing response. Recently, however, Premack (1963) obtained evidence suggesting that asymptotic reinforced response rate varies directly with the base rate of the instrumental response. The purpose of the present experiment was to inquire further into this relation.

An adequate test of this problem requires a procedure under which an extensive range of the independent variable (i.e., instrumental response base lines) may be produced. Differential weighting of the bar, a tradi-

[1] This report is based on a dissertation done under the direction of David Premack and presented in partial fulfillment of the requirements for the PhD degree, Department of Psychology, University of Missouri, 1962. This research was supported in part by Grant G-19574 from the National Science Foundation and by Grants M-5798 and MH-08775 from the National Institutes of Health.

[2] Now at Florida State University.

tional method of varying the base rate of bar pressing, is procedurally unsatisfactory because of the very limited range of responding that may be produced in this manner. A satisfactory method for investigating this problem was provided, however, by a reinforcement situation originally suggested by Premack (1962). In the present experiment rats were placed into an activity wheel into which a drinking tube extended. The wheel was locked and *S*s were required to lick in order to have the opportunity to run. The base rates of licking were manipulated by varying the contents of the drinking tube, while base rate of running, the reinforcing response, was held constant. This arrangement proved ideal for testing the problem; the precontingent probability of running was greater than that of base-line drinking for all tube contents, thus insuring reinforcement in the contingent situation, while the probability of base-line drinking varied substantially as a function of tube content, thus providing the desired wide range of the independent variable.

Method

Experimental Design

The *S*s were first given a period of wheel adaptation, then base-rate measures were taken of running alone, water drinking alone, and running and water drinking together. The *S*s were then trained to lick the water tube to run, then extinguished. The *S*s were then matched on the preceding measures, assigned to one of four experimental groups, and tested on the appropriate tube contents for 5 days. Finally, the main experimental design, involving a series of paired base-rate sessions preceding each FR schedule, was instituted.

Subjects

The *S*s were 16 female Sprague-Dawley rats, about 100 days old. They were individually housed, and maintained on ad-lib Purina mash and water.

Apparatus

Four modified Wahmann activity wheels that have been described in detail elsewhere (Premack & Schaeffer, 1962) were used for wheel adaptation. The apparatus used in the experiment proper has also been described in detail elsewhere (Premack, Schaeffer, & Hundt, 1964; Schaeffer & Premack, 1961). Briefly, this device consisted of a Wahmann activity wheel and drum assembly mounted in an insulated ice box hull, and a retractable drinking tube assembly. This test chamber permitted successive and simultaneous recording of running and/or drinking, and the programing of contingencies between drinking and running. Each one-quarter wheel turn and each lick were recorded on electric counters and an Esterline-Angus recorder.

Procedure

Wheel adaptation.—The *S*s were given 14 days of wheel adaptation in the activity wheels, to permit running to stabilize (Premack & Schaeffer, 1962, 1963). During this period *S*s were run for a 1-hr. A.M. and a 1-hr. P.M. session daily; thereafter, for all conditions they were run twice daily, in the A.M. and the P.M., for a 10-min. session.

Base rates and preconditioning.—Base rates were obtained, in the following order, of running alone, water drinking alone, and running and water drinking together. Next, all *S*s were conditioned for 5 days on FR 5 to lick the water tube to free the wheel for 10 sec. The *S*s were then given 1 day of extinction, primarily to extinguish nonlicking behaviors (Premack, 1962).

The *S*s were then matched and divided into four equal groups. Each group received only one of four different tube contents throughout the experiment. The tube contents were: empty tube (MT), tap water, 4% sucrose by weight, and 32% sucrose by weight. Matching of the *S*s was accomplished by using the preceding base-line and preconditioning measures. Mean number of wheel turns averaged over the last six sessions of the base-line period in which running and water licking were simultaneously available to *S*s were: 454.8, 429.6, 502.4, and 442.3, for Groups MT, water, 4%, and 32%, respectively. Mean number of instrumental licks averaged over the last two preconditioning sessions were: 271.2, 264.7, 253.2, and 255.7 for Groups MT, water, 4%, and 32%, respectively. Analyses of variance made on both sets of data failed to approach the 10% level; thus, the groups did not differ either in base

rate of running or in their general conditionability for the lick-to-run contingency.

Conditioning series.—After being assigned to one of the four main groups, each *S* was presented with the appropriate tube content for 5 days. These sessions provided base-rate measures of drinking for each of the tube contents. Thereafter, all four groups were subjected to the main experimental conditions which are summarized in Table 1. As may be seen in Table 1, the experiment consisted of three conditioning series: an FR 5, an FR 41, a return to FR 5, and finally, extinction. In each conditioning series, completing the FR freed the wheel for 10 sec. Preceding each conditioning series were a number of paired-response base-line sessions, i.e., sessions in which both the wheel and the tube were simultaneously available so that *S* could freely choose between them. The purpose of these paired base-line sessions was twofold: (*a*) they provided a recurrent check on the equality of base-rate running among the four groups, and (*b*) they provided a recurrent measure of base-rate licking from which increments in instrumental licking in the contingent situation could be evaluated.

With the exception of extinction, each stage of the experiment was carried to asymptote. Asymptote was considered attained when the submeans of licking in the first three and the last three sessions of six consecutive sessions differed by less than 10% from the grand mean of the six sessions. In addition, analyses of variance of licking in the A.M. and P.M. sessions of the last 3 days of all conditions preceding extinction indicated that there were no significant within-groups differences in licking between sessions. All of the data presented in this report are based entirely upon asymptotic responding in the last six sessions of each experimental condition.

TABLE 1

SEQUENCE OF EXPERIMENTAL CONDITIONS AND NUMBER OF SESSIONS PER CONDITION

Conditions	Number of Sessions
First paired base line (i.e., running and drinking simultaneously and freely available to *S*)	16
FR 5, lick-to-run contingency	48
Second paired base line	20
FR 41, lick-to-run contingency	20
Third paired base line	20
FR 5 replication, lick-to-run contingency	12
Extinction	6

RESULTS

Contingent licking.—Asymptotic reinforced instrumental licking as a function of base-line licking in the first paired base-line condition is shown in log-log coordinates in the lower panel of Fig. 1. The typical effects of reinforcement are seen clearly in the parameter, FR requirement, in this figure; licking for each group in the FR 41 exceeded licking for that group in either FR 5, and extinction produced higher licking rates for each group than any FR. Both contingent and extinction lick rate appear to be increasing monotonic functions of base-rate licking. Analyses of variance (Treatments $\times$ Sessions $\times$ *Ss*) of total licking in the last six sessions indicated that the trend seen in the licking curves was significant at less than the .005 level. The F values obtained for the FR 5, the FR 41, the FR 5 replication, and extinction, respectively, were 13.07, 20.50, 15.45, and 39.16. In all cases, $df = 3/12$. No significant sessions or interaction effects were found in any analysis.

Although the results are not shown here, replotting each of the licking curves seen in Fig. 1 as a function of base rate of licking in the immediately preceding paired base-line condition did not alter the increasing monotonic relation seen in Fig. 1.

Reinforced increments in licking.—Although the effects of reinforcement have traditionally been evaluated by examination of asymptotic reinforced instrumental response rate, a more suitable method, when base-line measures are employed, is to examine the

reinforced *increment* in contingent responding (i.e., asymptotic reinforced instrumental response rate minus the base rate of the instrumental response). As was shown in Table 1, paired response base-line sessions preceded each FR. Although the original base rates of the reinforcing response, running, were recovered in the subsequent paired base-line sessions, the original base rates of the instrumental response, licking, were not. By evaluating the reinforced increments in contingent licking (i.e., asymptotic reinforced licking minus base-rate licking in the paired base-line condition immediately preceding the reinforcement condition), these changes

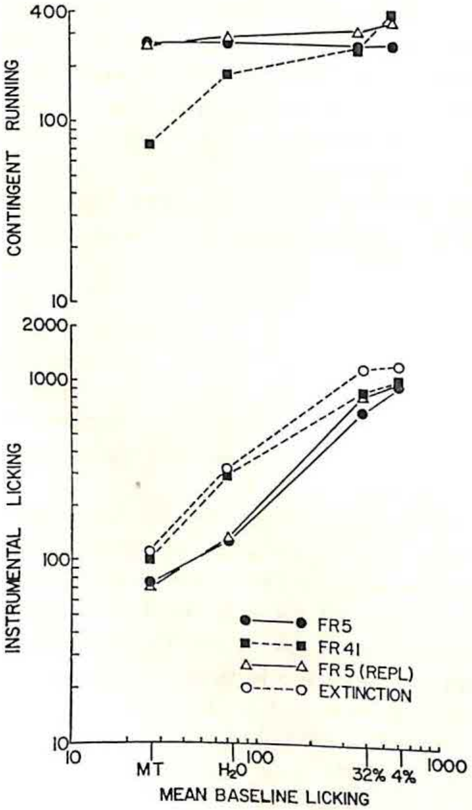


FIG. 1. Mean instrumental licks per session as a function of mean licks per session in the first paired base-line condition. (The upper panel shows mean contingent running per session in each FR.)

in base-line licking were taken into account.

The reinforced *increments* in instrumental licking as a function of licking in the paired base-line condition immediately preceding each FR and the extinction series are shown in log-log coordinates in Fig. 2. Each point in the curves represents the mean of the differences between instrumental licking and base-line licking in the paired base-line seesion which immediately preceded each FR series and the extinction condition. As can be seen, the magnitude of reinforced increments in instrumental licking increases with base rate of licking up to a point, after which they either level off or show a slight decline. More data points are needed from base lines higher than those of the present study to determine if the function is actually monotonic up to asymptote, or whether the relation is actually curvilinear. Separate analyses of variance of the curves of least and greatest slopes, the FR 5 replication and extinction cases, respectively, indicated that the trend seen in the former was not significant, $F\ (3, 12) = 1.44$, $p < .20$, whereas that for extinction was significant, $F\ (3, 12) = 6.54$, $p < .01$.

Of further interest was the finding that *replotting* the reinforced increments in licking in each contingency and extinction as a function of base-rate licking in the *first* paired base-line condition did not produce curves substantially different from those seen in Fig. 2. Although the absolute magnitude of the increments found with this method differed from those seen in Fig. 2, the shape of the functions obtained did not.

Contingent running.—The mean rates of the reinforcing running response in each FR contingency are shown in the upper panel of Fig. 1.

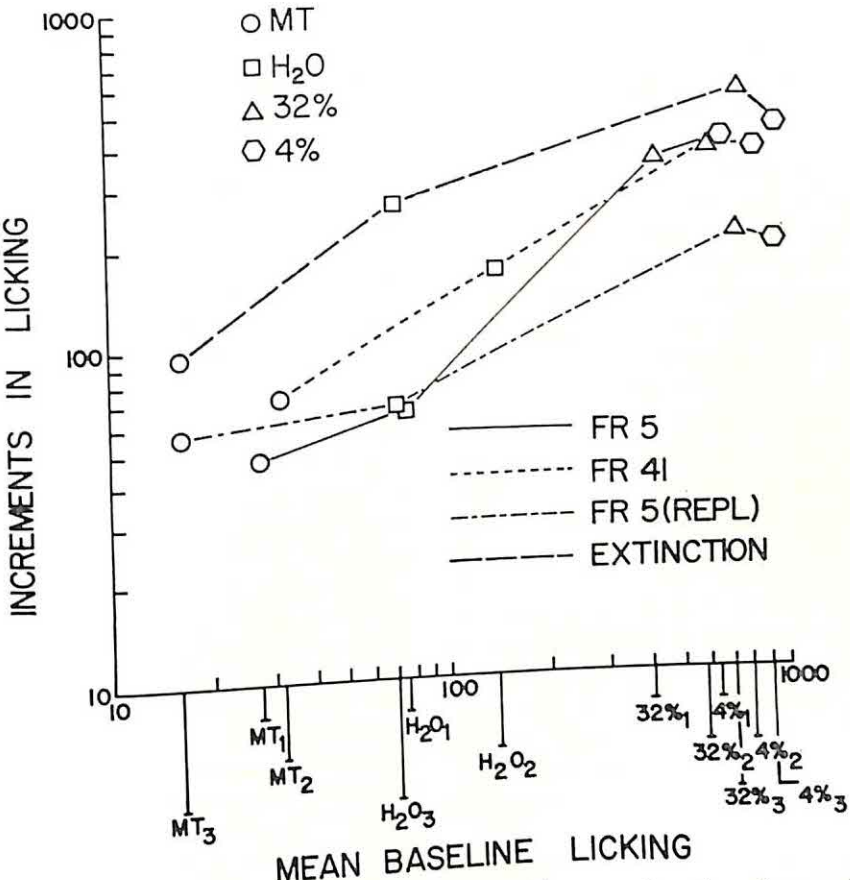


FIG. 2. Reinforced increments in instrumental licking as a function of mean licks per session in the paired base-line condition immediately preceding each contingency. (The subscripts 1, 2, and 3 by each group on the x axis refer to the first, second, and third paired base-line conditions, respectively. Thus, the FR 5 curve is plotted as a function of licking in the first paired base line, the FR 41 curve is plotted as a function of licking in the second paired base line, and the FR 5 replication and extinction curves are plotted as a function of licking in the third paired base line.)

As can be seen, there was a numerical difference between the groups on contingent running associated with all three FR conditions, with contingent running varying directly with instrumental licking. Mean number of runs per session for the empty tube, water, 32%, and 4% groups, respectively, for the FR 5 and FR 5 replication, respectively, were 269.9, 276.3, 278.9, 280.6; and 268.9, 291.6, 328.0, and 336.9. Mean number of runs per session in the FR 41 were 76.7, 188.4, 268.9, and 425.7. Only in the FR 41 did these differences in contingent running approach statistical significance, $F(3, 12) = 3.28$, $p < .10$.

The consistent differences between groups on contingent running might be due to differences in amount run per opportunity to run; or, to differences in momentary rate of running; or, to differences in number of opportunities to run. In order to establish the locus of these differences, Esterline-Angus records from the last two sessions of the FR 5 and FR 41 were examined. The number of trials

on which the FR was completed and followed by running, and the number of runs per trial were tabulated. Analyses of variance of mean runs per trial in the first 5, and last 5 min. of these sessions indicated that the groups did not differ in amount run per opportunity to run. Furthermore, the groups did not differ on momentary rate of running; over 80% of all completed wheel turns occupied 1 sec. or less. The groups did differ, however, on number of opportunities to run. Mean number of trials on which the FR was completed and followed by running for the empty tube, water, 32%, and 4% groups, respectively, for the FR 5 and FR 41, respectively, were 10.7, 11.1, 12.4, 12.7; and 2.9, 6.7, 11.9, 20.2. Of interest was the fact that number of opportunities to run varied directly with instrumental licking, indicating that in the contingent case the occurrence of the reinforcing response was itself partially dependent upon instrumental response occurrence.

In summary, asymptotic reinforced instrumental licking varied directly with base-rate licking. Reinforced *increments* in instrumental licking (reinforced licking minus base-line licking) appeared to be possibly a nonmonotonic function of base-line licking, with intermediate base rates resulting in the maximum increment, and smaller increments associated with both lesser and greater base rates of licking. The groups did not differ on base rates of running; the original base rates of running were recovered in the subsequent paired base-line sessions. The groups did differ, however, on rate of contingent running in all three FR cases, although the differences approached statistical significance only in the FR 41. Contingent running was found to vary directly with instrumental licking, and was strictly a function of the number of opportunities to run, there being no differences between the groups in either momentary rate of running or amount run per opportunity to run.

Discussion

The contention (Premack, 1963) that variations in the precontingent rate of the instrumental response influence the outcome of the reinforcement relation is borne out by the results of the present study. The effects of variations in instrumental response base rate are reflected not only in asymptotic reinforced responding, but in the magnitude of the reinforcement increment, and in the number of times the reinforcing response may occur in the contingent situation.

The stability of the relation between base-rate instrumental responding and reinforced responding is attested by the fact that it was found to hold for several FRs and for extinction. The possible generality of this finding is suggested by apparently comparable data obtained in situations differing markedly from that employed in the present study. Premack (1963) found a comparable relationship in monkeys, with preferred and nonpreferred manipulation responses as the reinforcing and instrumental responses, respectively. Asdourian (1962) has also reported what appears to be a comparable effect; septal stimulation in rats failed to reinforce licking water, but did reinforce licking a sucrose solution. With respect to the results of the present study, Asdourian's results may be attributable to the fact that the base rate of drinking a sucrose solution is considerably greater than that for drinking water. The results of both preceding studies, like those of the present study, indicate that the reinforcement increment is not determined solely by the reinforcer, but rather, is determined both by the reinforcer and the base rate of the instrumental response.

References

Asdourian, D. Interaction effects of intracranial stimulation with rewarding and aversive solutions. *J. comp. physiol. Psychol.*, 1962, **55**, 685–690.

Meehl, P. E. On the circularity of the law of effect. *Psychol. Bull.*, 1950, **47**, 52–75.

Miller, N. E., & Dollard, J. C. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

Premack, D. Reversibility of the reinforcement relation. *Science*, 1962, **136**, 255–257.

Premack, D. Rate-differential reinforcement in monkey manipulation. *J. exp. Anal. Behav.*, 1963, **6**, 81–94.

Premack, D., & Schaeffer, R. W. Distributional properties of operant-level locomotion in the rat. *J. exp. Anal. Behav.*, 1962, **5**, 89–95.

Premack, D., & Schaeffer, R. W. Some parameters affecting the distributional properties of operant level running in rats. *J. exp. Anal. Behav.*, 1963, **6**, 473–475.

Premack, D., Schaeffer, R. W., & Hundt, A. Reinforcement of drinking by running: Effect of fixed ratio and reinforcement time. *J. exp. Anal. Behav.*, 1964, **7**, 91–96.

Schaeffer, R. W., & Premack, D. Licking rates in infant albino rats. *Science*, 1961, **134**, 1980–1981.

(Received February 24, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 426–432

DELAYED COLD-INDUCED VASODILATATION AND BEHAVIOR[1]

WARREN H. TEICHNER

University of Massachusetts

The latency of cold-induced vasodilatation of the hand was found to be sensitive to the threat of shock and related to individual differences in performance on a task involving a conflict between a gain in money and a risk of shock.

When the body is kept warm and the hand is immersed in cold air or water, there is an immediate vasoconstriction in the fingers which may be observed as a rapid drop of surface temperature of the fingers. In 1929, Lewis observed that with continued immersion, a reflexive vasodilatation occurs. That is, after some time, the veins dilate bringing warm blood to the surface. This may be observed as a sudden reversal of the decreasing finger-temperature curve. Over a period of prolonged immersion, there may occur a phasic vasoconstriction-vasodilatation in the finger which is exhibited as a sinusoidal-like finger-temperature curve. This phenomenon, known variously as a *Lewis wave*, as a *reflexive vasodilatation*, as *cold-induced vasodilatation*, or as a *temperature-hunting reaction*, has been of considerable interest to physiologists concerned with cold acclimatization. Under standard testing conditions the parameters of the curve have been found to be fairly consistent from day to day for a given individual. However, between individuals there are large differences. Some people show the classical reflex described; others show a single vasodilatation to some upper limit of warming after which their hand cannot be cooled again; some rewarm after a brief immersion, while others require a relatively prolonged immersion. Finally, some people fail to show the hunting reaction at all and simply cool down to the level of the air or water. These individual differences have not yet been explained.

Meehan (1957) experimenting with daily hand cooling curves, reported an unusual result with one of his *S*s. On the first 2 days of immersion, *S*, a college student, provided identical 20-min. cooling curves. The curves were characterized by the usual negatively accelerated decreasing finger temperature and a sudden, marked vasodilatation at 10 min. of exposure. On the third day the curve was similar except that the amount of rewarming was somewhat less. On the fourth day Meehan reported that *S* arrived, having just completed an important examination. He appeared very tense and was visibly agitated. Nevertheless, Meehan conducted the cooling test as usual. On this day *S* showed no vasodilatation whatsoever. The cooling curve dropped monotonically to an asymptote at the temperature of the water. Furthermore, although the experience is normally reported as painful, *S* experienced so much pain on this day that he refused to return for further tests.

[1] This research was supported by the National Institute of Mental Health Grant NIH M4201.

Edna Dahlquist, Terrence Walsh, and Edward Youngling acted as research assistants.

The cold-pressor test has been available for a considerable time as an index of a hyperreactive vasomotor system and has been studied with ambiguous results in connection with emotional states. Meehan's observation goes considerably beyond the implications of the cold-pressor test, however. If verified, it has implications for the problem of individual differences in cold acclimatization and suggests an emotional basis for these differences. It also has implications for the psychophysiology of arousal. That is, it is conceivable that differences in the parameters of the cooling curve may be characteristic of emotional or arousal states. This study was initiated, therefore, to explore the possibility of relationships between Lewis waves and behavioral measures.

Experiment I

Method

Subjects.—Fifty-four undergraduate college students, male and female in equal number, enrolled in the summer session, were used as *S*s. The *S*s were paid for their participation.

Apparatus.—Since the parameters of the hand-cooling curve depend importantly on the body's thermal balance and since the hunting reaction is elicited most easily when the body is warmer than usual, the experiment was conducted in a temperature-controlled chamber maintained at 90° F., 50% RH. The water bath, controlled at a temperature of 34° F., was stirred slowly and automatically. The *S* sat with his right arm hanging loosely at his side and, during immersion, with his hand through a port into the water up to the wrist. No contact was made with the sides of the water bath by the hand nor of the chair with any part of the arm (so as not to influence circulation). The end of a 24-gauge copper-constantan thermocouple taped to the index finger provided a continuous temperature curve on a Leeds and Northrup "Speedomax." An external switching device permitted measurement of water temperature from a second thermocouple taped to the same finger so as to protrude .5 in. beyond the nail. Electric shock was provided with an inductorium and electrodes placed on the ventral and dorsal sides of the right arm just below the elbow. Shock levels, preselected by *E* as strong and very weak were calibrated by measuring the output of the inductorium at different positions through a 151,000-ohm load. Through this load a strong shock was 1,660 v., peak current of 11 ma.; weak shock was 242 v., peak current of 2 ma. Shock duration, constant at 2.0 sec. was controlled with an electronic interval timer.

Procedures.—To approach face validity in the sense of a genuine stress, *S*s were told that they were participating in the evaluation of a harmless, pain-killing drug. They were provided with a 5-point rating scale with which they were to report intensities of pain when they felt a change in pain status. They were also told that the cold immersion was a standard pain test and that in addition to evaluating the drug in regard to it, it was desired to evaluate the drug's effectiveness in reducing the pain of electric shock.

The experiment was carried out in four successive, daily, test periods. Each *S* was always run at the same time of day. Each test period required *S* to sit quietly for 20 min. prior to immersion in order to raise body warmth level. At a verbal signal he immersed his hand. The *S* was in isolation in the chamber, but could make his verbal reports or ask to be released with an intercom which carried his voice to *E* who watched him through a one-way window.

The *S*s were assigned at random into a strong-shock group, a weak-shock group, and a no-shock group. The two shock groups received shocks on the second and third experimental days 15 min. after immersion of the hand into the water. On the second and third days, and for the no-shock group on all days, the electrodes were not attached. A "drug" consisting of a small measure of colored water was taken orally on the second and third days. The *S*s knew every detail of the experimental procedures except when the shock would be administered during immersion and that they were being given the same substance as drug on the second and third days. They were told that a drug and a placebo were being administered in balanced fashion on those 2 days but not told when each was given.[2]

Results

Of a total of 54 starting *S*s, 13 were not completed. Of these, 7 were lost

[2] After completion of the experiment, *S*s were informed of the actual nature of the experiment.

before any immersion and 6 quit during or after the first immersion. These last 6 *S*s will be described briefly below.

Since it was desired to classify *S*s according to the nature of their cooling curves, but independent of the experimental treatments, an examination was made of the distributions of the parameters of the curve obtained on the first day. Among the parameters examined were the latency of vasodilatation, the rate of cooling to vasodilatation, the minimum finger temperature, maximum rewarming, rate of rewarming, and length of cycle. A vasodilatation was defined as a sudden warming of the finger of at least 3° F. A number of correlations were suggested, but the measures which served to discriminate among *S*s most clearly were the latency of vasodilatation and the cooling rate to vasodilatation. This is in agreement with a similar analysis performed by Yoshimura and Iida (1950). Since latency can be obtained from the temperature records by inspection, it was used for classification.

Using latency to vasodilatation during the first exposure, the 41 completed *S*s were classified into three groups: (*a*) those with a latency of <270 sec., (*b*) those with a latency of 270–400 sec. inclusive, and (*c*) those with a latency of > 400 sec. The resultant distribution of *S*s in the experimental design is shown in Table 1.

Using latency of vasodilatation as

TABLE 1

NUMBER OF *S*S IN EXP. I

	< 270	270–400	> 400	*N*
No shock	4	8	6	18
Weak shock	2	6	5	13
Strong shock	4	5	1	10
N	10	19	12	41

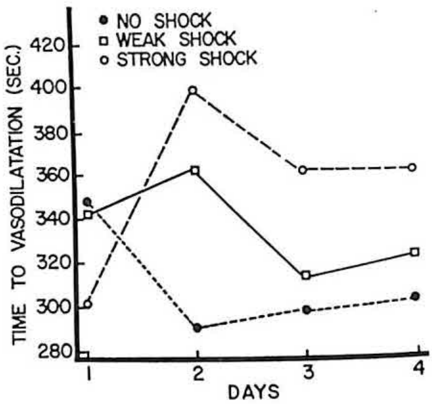


FIG. 1. Latency of vasodilatation of the 270–400 sec. group.

the dependent measure presented a logical problem. If the effect of the treatments were to reduce the latency, then the < 270-sec. group was restricted by already having a very short latency. Conversely, the > 400-sec. group contained *S*s who never showed a Lewis wave and if the effect of the treatments were to be an increase in latency, then this group could show no change. Thus, the clearest effect of the treatments was most likely to be seen in comparisons within the 270–400 sec. group. These data, therefore, are those to be presented.

Figure 1 presents the mean time to vasodilatation of the three treatment groups on each day. The figure shows that the duration of vasoconstriction was reduced in the control group after the first experience. In the weak-shock group it was increased on Day 2, but reduced thereafter although it never reached the latency of the no-shock group. The high-shock group started with a slightly shorter latency on Day 1 than the other two groups, but exhibited a marked prolongation of vasoconstriction on Day 2, an increase of about 100 sec. between days. The latency on the remaining 2 days was shorter than that of Day 2, but

still represented a great increase over Day 1 and over the other two groups.

Figure 2 presents the same results using rate of cooling to vasodilatation as the dependent variable. Since this measure takes account of both latency and temperature, it must be interpreted differently than either alone. That is, here a decreased rate represents an increased or prolonged vasoconstriction; a decreased or shortened vasoconstrictive effect is shown by an increased rate. As may be seen, using this measure which smoothed the data considerably, allows generalization of the latency results to cooling rate.

The pain reports elicited from *S*s were highly variable and difficult to interpret after the first day possibly because of the administration of shock after that day. The mean pain ratings obtained on the first 15 min. of the first day are shown in Fig. 3 and here it may be seen that reported pain decreased with continued immersion for all groups. The > 400-sec. group consistently reported greater pain than the other two groups. The other two groups experienced essentially the same amount of pain.

Earlier we noted the loss of six *S*s who quit during or after the first day's test. Thus, none of these *S*s experienced the electric shock. All of them

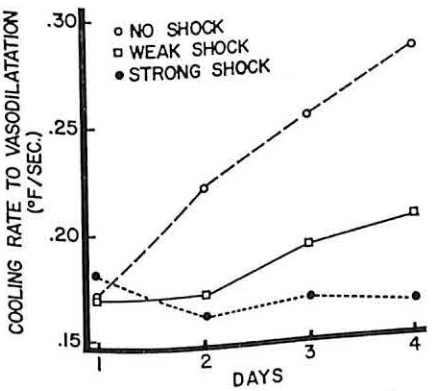


FIG. 2. Cooling rate to vasodilatation of the 270–400 sec. group.

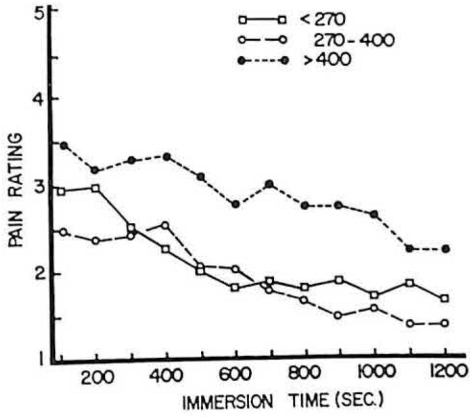


FIG. 3. Mean pain ratings of the three Lewis wave groups as a function of immersion time.

were immersed for at least 400 sec. Every one was found to be characterized by three things: (*a*) a complete absence of vasodilatation during the first day's immersion, (*b*) a self-initiated withdrawal from the experiment, and (*c*) a current or recently experienced intense, real life, emotional experience.

EXPERIMENT II

The purpose of this experiment was to extend the previous one by examining the relationships between individuals classified by first exposure latency to vasodilatation and performance in an independent situation in which performance might be thought of as sensitive to arousal.

Method

A *conflict-uncertainty* task was developed using preliminary experimentation to select its procedural and apparatus characteristics. The apparatus was a small panel containing six doorbell buttons arranged in a circle. A seventh button in the center of the circle turned on a pilot light at the top of the panel. The *S* was instructed to turn on the light with the center button and to turn it off with one of the peripheral buttons. He was told that on the first trial all of the six buttons would turn the light off; on the second trial all but one

would work; on the third trial all but two, etc., until only one button worked. He was also told that the arrangement of wrong buttons would not be consistent from trial to trial. Actually, all buttons always turned off the light.

Just before the first trial S was given a strong, 2-sec. shock defined as in Exp. I, through electrodes attached to the wrist of the working hand. He was informed previously that this was a sample of the shock he would receive each time he pressed a wrong button. He was informed that he would accumulate a sum of money throughout the test regardless of errors, starting with 5 cents on the first trial, which it will be recalled had no risk, and cumulating to $1.00 if he completed all trials. He was told that he could quit at any time and keep his accumulated earnings. Since, although unknown to S, all buttons were correct on all trials, no S ever received more than the sample shock. The first trial began 60 sec. after the shock. In addition to the procedures described, continuous recordings were made of the index-finger temperature of the nonworking hand.

The hand-cooling procedures were carried out as before with a 20-min. equilibration to 90° F., 50% RH, followed by a 20-min. water immersion at 34° F. No shock or electrodes were involved.

Subjects.—Thirty-three male, undergraduate students were used as paid *Ss*. None of these had had any prior experience with any of the procedures used.

Results

Nine *Ss* vasodilated between 270 to 400 sec.; 24 vasodilated at > 400 sec.; none vasodilated at < 270 sec. Although most of the 33 *Ss* required only a few minutes to complete the task, one S of the > 400-sec. group, required over an hour. This S's data will be presented separately. The remaining *Ss* are shown in Fig. 4 in terms of mean time to select among the six buttons and mean finger temperature of the nonworking hand. The finger temperatures shown are for 30 sec. before the sample shock, 30 sec. after it, and at the time of each decision.

The results shown in the upper part of Fig. 4 indicate that the > 400-sec.

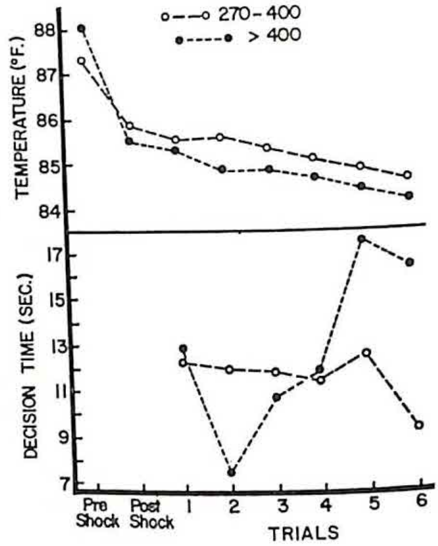


FIG. 4. Finger temperature (top) and decision time (bottom) of the two Lewis wave groups during the conflict-uncertainty task.

group started with a higher finger temperature. With shock the > 400-sec. group dropped to and remained at a lower finger temperature than the other group which indicates a greater rate and amount of vasoconstriction. Both groups showed a progressively increasing vasoconstriction as the task progressed.

Inspection of the decision times in the lower half of the figure shows that the > 400-sec. group started with a slightly longer decision time than did the other group, but with the first involvement of risk (Trial 2), this group showed a marked decrease in decision time. After the second trial, the curve rises steeply so that at first it still represents a faster decision time than for the 270–400 sec. group, but later it crosses the curve of this group and as the risk increases further, it rises to a maximum of over 17 sec. and then drops slightly. The decision times of the 270–400 sec. group, on the other hand, showed a more-or-less continuous decrease with increasing risk, and on the last trial, not unlike

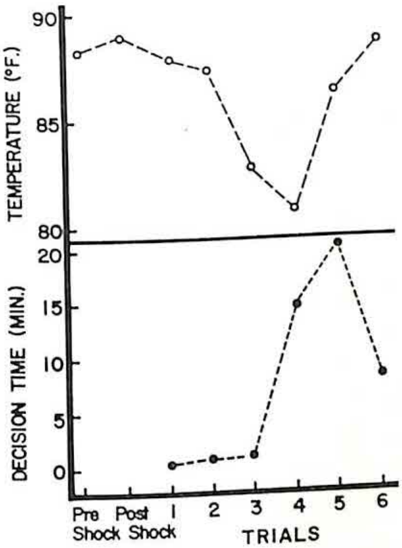


FIG. 5. Finger temperature and decision time of one extreme S during the conflict-uncertainty task.

the > 400-sec. group, they showed an end effect.

One S in the last experiment, not included in the above results, required over an hour to complete the task. His data are shown in Fig. 5. These data are very interesting although not completely consistent with the previous ones.

Note that the ordinate of Fig. 5 is in minutes rather than seconds. The S started with a slow decision time. His decision times then increased linearly until the third trial. This represents a large increase relative to the previous figure. On the fourth trial it took him over 14 min. to make a decision, on the fifth trial over 20 min. Looking at the finger temperature of this S, there is some suggestion of a mirror-image relationship although with a time lag. In any case, it is clear that as finger temperature reversed, decision time showed a decrease, although with a lag. This S started with a finger temperature of 88.2° F. as compared to 88.0° F. for the mean of the > 400-sec. group in the last figure. However, the minimum temperature reached for that group was 84.4° F. whereas the minimum of this S was 80.6° F.

Discussion

The results of Exp. I appear very clear in supporting Meehan's (1957) observation that cold-induced vasodilatation may be influenced by emotional stress. That experiment also suggests that the latency of vasodilatation can be increased by the threat of shock and that the degree to which the latency is affected depends on the intensity of the threat. The continuation of effect to the fourth day by both shock groups, even though they were assured that they would not be shocked on that day, suggests a conditioning phenomenon, and, to the extent that this can be verified, suggests the possibility of a more or less lasting alteration of the cooling curve. On the other hand, the results suggest that in the absence of additional stress, Ss habituate to the cold water experience.

When the results of the treatment effects of Exp. I are considered along with the characteristics of the dropout Ss of that experiment and along with related observations (Meehan, 1957; Teichner, 1963), the hypothesis is suggested that ability to adapt (latency) to cold stress under conditions which permit a reflex vasodilatation is related to a relatively high chronic arousal characterized by a tendency to vasoconstrict under emotion-producing stresses. A direct source of support for this hypothesis comes from results showing a delayed indirect vasodilatation in schizophrenics as compared to normals (Henschel, Brozek, & Keys, 1951) and vasomotor habituation of the orienting response (Sokolov, 1960; Unger, 1964). Related support may also be found in other studies which have reported inadequate peripheral vascularization in schizophrenics (e.g., Abramson, Schloven, & Katzenstein, 1941; Shattock, 1950) and experiments finding vasoconstriction following induced emotional states (Hovland & Riesen, 1940;

haps it should be mentioned here that Broadbent's original findings, utilizing lists of digits, were replicated at Indiana University in 1962 using the same equipment as was used in the present experiment.) Thus, even if Broadbent's estimate of the time required to switch attention from channel to channel were revised, his theory as it now stands would not be able to account for the observed results of this experiment.

It is quite possible that telling half the *S*s in the present experiment that if they switched channels in a specified way some of their answers might make sentences may have given them some sort of set which enabled them to listen to the words all at once in some way and then rearrange them to get sentences before responding. In fact, many of the *S*s did manage to rearrange a number of the SCR lists and get sentences out of them. In this regard it is interesting to note that after the experiment was over several of the *S*s who had received the "channel" instructions reported spontaneously that some of the lists made sentences even though they should not have been switching channels at all. Again this indicates that these *S*s were attending in a manner which Broadbent's theory says they should be unable to do, at least at the rate of presentation used in the present experiment.

REFERENCES

Broadbent, D. E. The role of auditory localization in attention and memory span. *J. exp. Psychol.*, 1954, **47**, 191–196.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Moray, N. Broadbent's filter theory: Postulate H and the problem of switching time. *Quart. J. exp. Psychol.*, 1960, **12**, 214–220.

(Received February 27, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 437–438

SUPPLEMENTARY REPORT

EFFECTS OF CS AND UCS CHANGE ON EXTINCTION OF THE CONDITIONED EYELID RESPONSE

LOUIS E. PRICE, DAVID W. ABBOTT,[1] AND WILLIAM E. VANDAMENT[2]

University of Massachusetts

Ss were given 20 extinction trials, following 70 acquisition trials on continuous reinforcement, under 1 of 6 combinations of CS and UCS conditions. The CS was unchanged from acquisition, shortened or extended, and the UCS was delayed or omitted. The CS conditions did not have a significant effect on extinction while the delayed UCS led to slower extinction than the omission condition. There was no CS-UCS interaction.

McAllister (1953) and Reynolds (1958) have both reported that CRs extinguish faster under a UCS omission procedure than with the technique of extending the CS-UCS interval to 2,500 msec. following a partial-reinforcement schedule during acquisition.

Spence, Rutledge, and Talbott (1963) and Reynolds (1958) found that following continuous reinforcement there were no significant differences in extinction rate with the UCS delay vs. UCS omission procedures.

Spence et al. discuss these results in terms of a discrimination hypothesis. They point out that the ability of *S*s to discriminate the change from acquisition to extinction is a function of the degree of stimulus change from acquisition to extinction. Following this interpretation, the UCS delay procedure should lead to slower extinction since the UCS omission would provide a major stimulus change from acquisition to extinction. Following a partial-reinforcement schedule, the UCS delay in extinction should be an even less discriminable change since in nonreinforced trials during acquisition *S*s would have experienced this same UCS condition. Following 100% reinforcement in acquisition, it would appear that both UCS omissions and UCS delay are discriminable enough changes to lead to rapid extinction since neither condition had been present during acquisition.

Spence (1963) points out another stimulus variable which may affect the discriminability of the change from acquisition to extinction. This is the duration of the CS. When the UCS is delayed in extinction, the CS duration is increased from its acquisition interval to overlap with the UCS. He suggests that this factor following continuous reinforcement for the UCS delay extinction may be as effective a cue for *S*s that things have changed as the UCS omission is. In support of this notion, he reports that UCS delay does lead to slower extinction than UCS omission following continuous reinforcement if the CS is of long duration during both acquisition and extinction, thus eliminating the cue of CS change.

The present study was designed to further evaluate the effects of CS and UCS change from acquisition to extinction following continuous reinforcement. Three CS conditions were combined factorially with two UCS conditions. The CS was shortened, lengthened, or unchanged in going from acquisition to extinction while the UCS was either delayed or omitted in extinction.

Method.—Forty-eight men and 48 women from psychology courses at the University of Massachusetts served as *S*s. In addition, one man was discarded from the study due to apparatus failure.

A dental chair for *S*s was located in a semidarkened room adjoining the room containing the recording apparatus and stimulus controls. The CS was located 50 in. in front of *S* and consisted of an increase in the illumination of a 6-cm. wide circular milk-glass disk from 0.032 to 1.364 apparent foot-candles (ftc.). The UCS was a corneal air puff of 2-psi intensity. Background noise from an electric fan prevented *S*s from hearing any apparatus noise. The intertrial interval was controlled by a Gerbrands programmer while CS duration, CS-UCS interval, and UCS duration were controlled by a series of Hunter timers. A record of eyelid movement, via a microtorque potentiometer, an amplifier, and

[1] Now at Stetson University, Deland, Florida.
[2] Now at Harpur College, Binghamton, New York.

TABLE 1

MEAN NUMBER OF CRs DURING 10-TRIAL BLOCKS OF EXTINCTION TRIALS

Trial Blocks	CS					
	Shortened		Unchanged		Lengthened	
	Omitted UCS	Delayed UCS	Omitted UCS	Delayed UCS	Omitted UCS	Delayed UCS
1	1.94	5.06	3.81	5.25	1.94	5.13
2	.88	2.88	1.88	3.63	.69	3.81

pen motor, and a record of stimulus onset were obtained with a pen recording system using a paper speed of 120 mm. per sec.

Once *S* was seated in the dental chair and the recording apparatus had been attached, he was read the standard instructions which were designed to create a relaxed, passive attitude. Each *S* was then given 70 acquisition trials and, with no delay other than the usual 15-sec. intertrial interval, 20 extinction trials. During acquisition the CS duration was 550 msec. with a CS-UCS interval of 500 msec. and a UCS duration of 50 msec. For the extinction trials *S*s were randomly assigned to one of the six extinction conditions, within the restriction that half of each group be males and half be females. The extinction groups were orthogonal combinations of three CS conditions and two UCS conditions. The UCS was delayed to 2,500 msec. or omitted while the CS duration remained unchanged at 550 msec., was shortened to 350 msec., or was extended to 2,550 msec. Responses falling in the 200–500 msec. range from CS onset, with a recorded deflection of 1 mm. or more, were counted as CRs during both acquisition and extinction.

Results and discussion.—All groups were responding at about 84% CRs during the last block of 10 acquisition trials.[3] A factorial analysis of variance was carried out on the mean number of CRs for the last 10 acquisition trials with CS duration and UCS conditions as variables to check that groups had reached the same level of performance prior to extinction. This analysis produced no significant *F*s for either main effects or the interaction. Table 1 presents the mean number of CRs for each 10-trial block of extinction trials for each combination of CS and UCS conditions. An analysis of variance was performed on these data with CS and UCS conditions and sex as between-*S*s variables and trials as a within-*S*s variable. The two significant *F*s were Trials, $F\ (1, 84) = 80.41$, $p < .001$, and UCS Condition, $F\ (1, 84) = 9.45$, $p < .01$.

The significant UCS condition was due to a slower extinction rate for the delayed UCS procedure. This result, along with the finding that the CS condition was not significant ($F < 1$), suggests that the possible *D* level effect of omitting the UCS, as pointed out by McAllister (1953), may also be a relevant variable here.

It is clear from our results and those of previous investigations that further work needs to be done to understand the effects of variables which influence the rate of extinction in the eyelid situation.

[3] Analyses reported here were performed on the data from all 96 *S*s. Twenty of these were classified as voluntary-form responders on the basis of the form of their anticipatory eyelid response (Spence & Ross, 1959). Separate analyses of both acquisition and extinction data were made excluding the 20 voluntaries, however, the results did not differ from what is reported in this article.

REFERENCES

MCALLISTER, W. R. The effect on eyelid conditioning of shifting the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 423–428.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

SPENCE, K. W. Cognitive factors in the extinction of the conditioned eyelid response in humans. *Science*, 1963, **140**, 1224–1225.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

SPENCE, K. W., RUTLEDGE, E. F., & TALBOTT, J. H. Effects of number of acquisition trials and the presence or absence of the UCS on extinction of the eyelid CR. *J. exp. Psychol.*, 1963, **66**, 286–291.

(Received February 5, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 4, 439–440

Beginning with the first issue of Volume 69 and continuing through the remainder of that Volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are the entire backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication. The articles listed below are scheduled to appear approximately 12 months hence.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Averaging versus Adding as a Stimulus Combination Rule in Impression Formation: Norman H. Anderson*: Department of Psychology, University of California, Los Angeles, California 90024.

Transfer of a Pattern-versus-Component Discrimination Following Training in a Probabilistic Situation: Marcia D. Johns*: Department of Psychology, Colorado College, Colorado Springs, Colorado 80903.

Role of Reward Magnitude and Incomplete Reduction of Reward Magnitude in the Frustration Effect: James H. McHose* and H. Wayne Ludvigson: Department of Psychology, Southern Illinois University, Carbondale, Illinois 62903

A Test for a Learned Drive Based on the Hunger Drive: John H. Wright*: Psychological Laboratory, Gilmer Hall, University of Virginia, Charlottesville, Virginia.

Internal Consistency of Subjective Probabilities: Cameron R. Peterson,* Z. J. Ulehla, Alan J. Miller, and Lyle E. Bourne, Jr.: Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan, P.O. Box 618, Ann Arbor, Michigan 48107.

Discriminated Avoidance Learning as a Function of Parameters of Discontinuous Shock: M. R. D'Amato,* Donald Keller, and Gerald Biederman: Psychology Department, Rutgers University, New Brunswick, New Jersey 08903.

Affect as a Function of Stimulus Variation: Paul C. Vitz*: Department of Psychology, Stanford University, Stanford, California 94305.

UCS Intensity and Avoidance Learning: Seymour Levine*: Department of Psychiatry, Stanford Medical Center, 300 Pasteur Drive, Palo Alto, California.

Effects of Stimulus Complexity and Restrictive Response Subclasses on Observer Responses: Irwin L. Goldstein*: Laboratory of Aviation Psychology, Ohio State University, 1314 Kinnear Road, Columbus 12, Ohio.

Interlimb and Interjoint Transfer of a Kinesthetic Spatial Aftereffect: G. Singer and R. H. Day*: Department of Psychology, University of Sydney, Sydney, N. S. W., Australia.

Dimensionality in Human Information Storage: Melvin H. Rudov*: 5628 Leibold Drive, Dayton, Ohio.

Conscious Mediating Processes in a Problem-Solving Task: Leonard S. Stein*: Department of Psychology, Duke University, Durham, North Carolina 27706.

Formation, Maintenance, Generalization, and Retention of Response Hierarchies: Albert E. Goss* and Nancy J. Cobb: Department of Psychology, University of Massachusetts, Amherst, Massachusetts.

Visual and Motor Components of an Experimentally Induced Position Preference in Multiple-Probability Learning: Stanford H. Simon*: VA Hospital, Tomah, Wisconsin 54660.

Effect of Overtraining on Reversal and Extradimensional Shifts: Thomas J. Tighe*: Department of Psychology, Dartmouth College, Hanover, New Hampshire.

Effects of Independent and Dependent Outcome Values upon Bets: Francis W. Irwin* and Joan Gay Snodgrass: Department of Psychology, 106 College Hall, University of Pennsylvania, Philadelphia, Pennsylvania 19104.

---

* Asterisk indicates author for whom address is supplied.

Serial Acquisition as a Function of Item Probability and Sequential Probability: James F. Voss*: Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

Intermodal Transfer in a Paired-Associates Learning Task: Gary L. Holmgren, Malcolm D. Arnoult,* and Winton H. Manning: Department of Psychology, Texas Christian University, Fort Worth, Texas 76129.

Short-Term Memory for Motor Responses: Jack A. Adams*: Aviation Psychology Laboratory, University of Illinois Airport, Savoy, Illinois.

Serial Learning as a Function of Meaningfulness and Mode of Presentation with Audio and Visual Stimuli of Equivalent Duration: Rudolph W. Schulz* and Richard A. Kasschau: Department of Psychology, University of Iowa, Iowa City, Iowa.

Phonemic Similarity and Interference in Short-Term Memory for Single Letters: Wayne A. Wickelgren*: Department of Psychology, E10-035, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Intertrial Stimuli and Generalization of the Conditioned Eyelid Response: John W. Moore* and Frederick L. Newman: Department of Psychology, University of Massachusetts, Amherst, Massachusetts.

Visual Search and Immediate Memory: Ira T. Kaplan,* Thomas Carvellas, and William Metlay: Dunlap and Associates, Inc., Darien, Connecticut 06821.

Temporal Conditioning of GSR: Russell A. Lockhart*: Department of Psychology, University of California, Santa Barbara, California 93106.

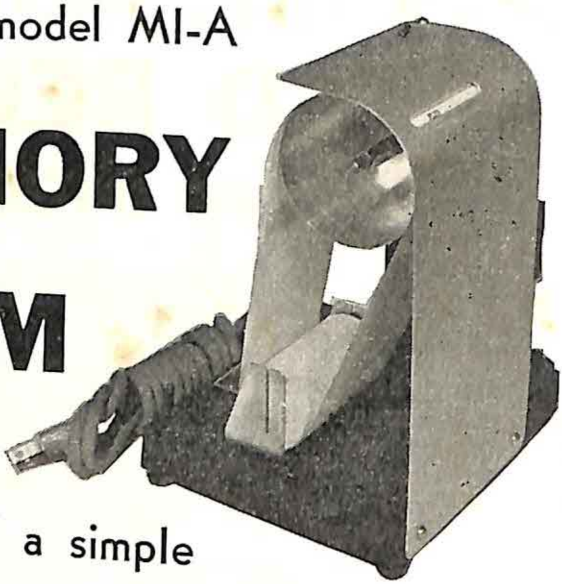
Information Capacity of Discrete Motor Responses under Different Cognitive Sets: Paul M. Fitts* and Barbara K. Radford: Department of Psychology, University of Michigan, Ann Arbor, Michigan.

Test of a Prediction of Stimulus Sampling Theory in Probability Learning: Norman H. Anderson*: Department of Psychology, University of California, Los Angeles, California 90024.

Effect of Spatial Parameters on the Vibrotactile Threshold: Ronald T. Verrillo*: Laboratory of Sensory Communication, Syracuse University, Syracuse 10, New York.

A Test of Underwood's Theory of Distributed Practice: John Jung*: Department of Psychology, California State College, Long Beach, California 90804.

Reexamination of Implicit Verbal Chaining: Burton H. Cohen* and Donald A. MacNeil: Psychology Department, Lafayette College, Easton, Pennsylvania.



Journal of Experimental Psychology

VOL. 69, No. 5 MAY 1965

STIMULUS DETERMINANTS OF CHOICE BEHAVIOR IN VISUAL PATTERN DISCRIMINATION [1]

JAQUES KASWAN, STEPHEN YOUNG, AND CHARLES Y. NAKAMURA

University of California, Los Angeles

5 Ss participated in a forced-choice, pattern-recognition study. In each of a total of 10 sessions, Ss were required to identify tachistoscopic presentations from a set of 3 or 2 patterns. In separate sessions each S received 1 set of 3 patterns and 3 subsets of 2 patterns, the latter containing all the possible combinations of the 3-alternative set. Exposure durations ranged from 4 to 128 msec. It was assumed that spatial distance characteristics of the patterns and exposure duration were stimulus dimensions which help determine the perception of these patterns. This assumption was used to predict (a) the order of response probabilities to each pattern at different levels of exposure duration, (b) the accuracy of pattern identification as a function of exposure duration, and (c) the response probabilities to each pattern in 2-alternative sets from the responses to these patterns in the 3-alternative set. The predictions were generally confirmed, indicating that these stimulus dimensions may contribute substantially to the determination of choice behavior.

The study of choice behavior in perceptual-recognition tasks has recently been a matter of considerable interest. In these tasks, *S* is generally required to specify which member of a set of alternative stimuli (e.g., A,B,C) has been presented by using response alternatives (a,b,c) that are in one to one correspondence with the stimuli. An adequate account of such choice behavior should include, first, predictions of the probability with which each response alternative occurs to each stimulus in a given set, and secondly should predict how the response probabilities obtained from a particular set of stimuli (e.g., the set A,B,C) are related to a subset of this set (e.g., the stimulus set A,B with response alternatives, a,b). In existing formulations, these predictions are based either on empirically obtained rules for choice behavior (Clarke, 1957), or on logical conceptions which assume that choice behavior results from a decision process applied to some internal representation of the relations between the stimuli within the set of alternatives (Luce, 1963). These formulations make no explicit

[1] This study was supported by a grant from the National Science Foundation (G-9589) and a fellowship (for Stephen Young) from the Public Health Service (5-F1-MH-19,664-02). We wish to thank Sally Haralson and Ruth Cline for help in collecting and analyzing the data.

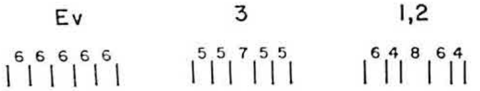


FIG. 1. Sample sections of patterns and distance between spaces in mm.

attempt to consider the specific role of physical stimulus dimensions as determinants of choice behavior. For example, consideration of stimulus effects is often limited to differences in discriminability between alternatives which might consequently be arranged on a similarity scale (Luce, 1963), or the number of alternatives and signal-to-noise ratios (e.g., Attneave, 1959; Garner, 1962).

Some of our previous work has shown that exposure duration and spatial arrangement of stimuli are important stimulus dimensions determining pattern perception (Kaswan, 1958; Kaswan & Young, 1963). A major goal of the present study is to explore the utility of these stimulus dimensions in predicting choice behavior in a tachistoscopic pattern-recognition task. The specification of these variables can be illustrated with reference to Fig. 1, which shows patterns constructed on the basis of formulations developed in our previous studies (Kaswan, 1958; Kaswan & Young, 1963). Each of these patterns can be defined in terms of a sequence of spaces, with each space delimited by a pair of vertical lines. Previous work suggests that beyond its role in detection, exposure duration determines the perception of existing differences among spaces within the pattern presented, according to the following assumptions: (*a*) If the pattern contains two adjacent spaces x and y, which are equal in size, these spaces should not be perceived as different in size, regardless of exposure duration. Thus, the Ev pattern in Fig. 1, should always be perceived as equally spaced. (*b*) If Space x and Space y differ in size, then the probability of perceiving them as different decreases with decreasing exposure duration. Thus, at very brief exposures, differences in the size of spaces within the "3" or "1,2" pattern (see Fig. 1), are unlikely to be perceived and, therefore, all spaces should be perceived as equal. (*c*) If Space x and Space y differ in size, the greater the difference, the greater the likelihood that they will be perceived as different at a given exposure duration. Thus, longer presentations of the 1,2 pattern should be required for the perception of the difference between the 4- and 6-mm. spacing than between the 4- and 8-mm. spacing within this pattern. These assumptions should be relevant for any stimuli where relative distance between elements determines groupings like those described in the Gestalt literature as following the principle of proximity.

The purpose of this study was (*a*) to test predictions derived from these assumptions for the relative frequency with which each response alternative would occur to each pattern as a function of exposure duration when all three patterns are stimulus-response alternatives, and (*b*) to use these assumptions to translate the results obtained in the three-pattern task into predictions of the exact probability with which each available response alternative occurs when only two of the patterns are possible stimulus-response alternatives (i.e., Ev-3; Ev-1,2; 1,2-3). The specific predictions will be discussed in the results section.

METHOD

Materials.—Sections of the three patterns used are shown in Fig. 1, together with their dimensions. They were constructed on the basis of assumptions which are outlined in other parts of this paper. The patterns were

drawn through the center of an 8-in. square field and multilithographed onto flat-white cardboard sheets to permit the arrangement of randomized decks of stimuli. The Ev, 3, and 1,2 patterns contained a total of 26, 27, and 24 black vertical lines and their total lengths on the stimulus card were $6\frac{1}{16}$, $6\frac{1}{16}$, and $5\frac{11}{16}$ in., respectively. Each of these patterns subtended a visual angle of approximately 11°.

Apparatus.—The patterns were presented with a two-field, electronically controlled mirror tachistoscope. This tachistoscope is described in more detail elsewhere (Kaswan & Young, 1963). The fixation field contained two 1-in. markers at each end of the field to indicate the horizontal line on which the stimulus appeared.

Procedure.—In each session, one of the possible combinations of the three patterns constituted the set of stimulus and response alternatives. Combinations used were the entire set of three stimuli (three-alternative condition) and each of the three possible subsets of two stimuli (two-alternative combinations). Each *S* received all of these combinations.

After a .5-sec. warning buzzer, *S* presented the stimulus to himself by pulling a lever as soon as he was ready. Each of the stimuli used in a given session was initially presented for two, .5-sec. exposures in order to inform *S* of the alternatives available in the session. The *S* was requested to describe these initial presentations in his own words. After correct descriptions were given, which was always the case, *S* was instructed to use the stimulus names "even," "three," and "one-two" to identify presentations which corresponded to the Ev, 3, 1,2 patterns, respectively. In addition, *S* was instructed to respond with the word "nothing" if he saw no sign of a stimulus.

After *S* was given the above response instructions, he received a practice series in which each stimulus was presented once at 10, 50, and 100 msec., in random pattern and duration order. Upon completion of practice, each stimulus was presented under each of eight exposure durations (4, 6, 8, 10, 16, 32, 64, and 128 msec.), 5 times in each three-alternative session and 10 times in each two-alternative session. Exposure duration and pattern sequence were randomized in each session for each *S*. There was an interval of about 11 sec. between presentations of each pattern.

Each *S* had 10 sessions in two successive blocks of 5 sessions each. A block consisted of 1 session of each two-alternative combination followed by 2 three-alternative sessions. The session sequence for the two-alternative combinations was randomized within blocks for each *S*. Thus, over the 10 sessions, each *S* received 20 replications of each pattern and exposure duration under each stimulus combination. These were completed within 4 wk. Six blank cards were also randomly inserted for each of the 20 exposure-duration-pattern replications given each *S*, to test for false positive responses.

Exposure luminance remained constant at about 11.84 mL. All viewing was binocular. The experiment was administered to individual *S*s in a dark room, except for a 6-w. light used by *E*.

Subjects.—The *S*s were one male senior and four male graduate students, all in psychology. No *S* was sophisticated in the area of perception; all had little or no experience in serving as *S*s in such experiments, and were naive as to the purpose of the study.

RESULTS

Figure 2 shows the percentage of each error response (solid and dashed lines) and the percent "nothing" responses (dotted lines) as a function of exposure duration for each pattern under all conditions for the five *S*s combined. Figures 2a, 2b, and 2c show the results for each pattern in the three-alternative condition. Figures 2d, 2e, and 2f show these results for the two-alternative combinations.

In order to utilize the most stable results, only the last 10 of the 20 replications for each pattern and exposure duration for every *S* and combination were used in the analyses presented below. Each point in Fig. 2 is thus based on 50 exposures. The error response percentages (solid and dashed lines) were calculated from the total number of responses excluding "nothing" responses.

The dotted lines in each part of Fig. 2 show that the percentage of "nothing" responses was about the same to each pattern in every combination. All blank card presentations were responded to with "nothing" indicating that false positive responses can be discounted in these results.

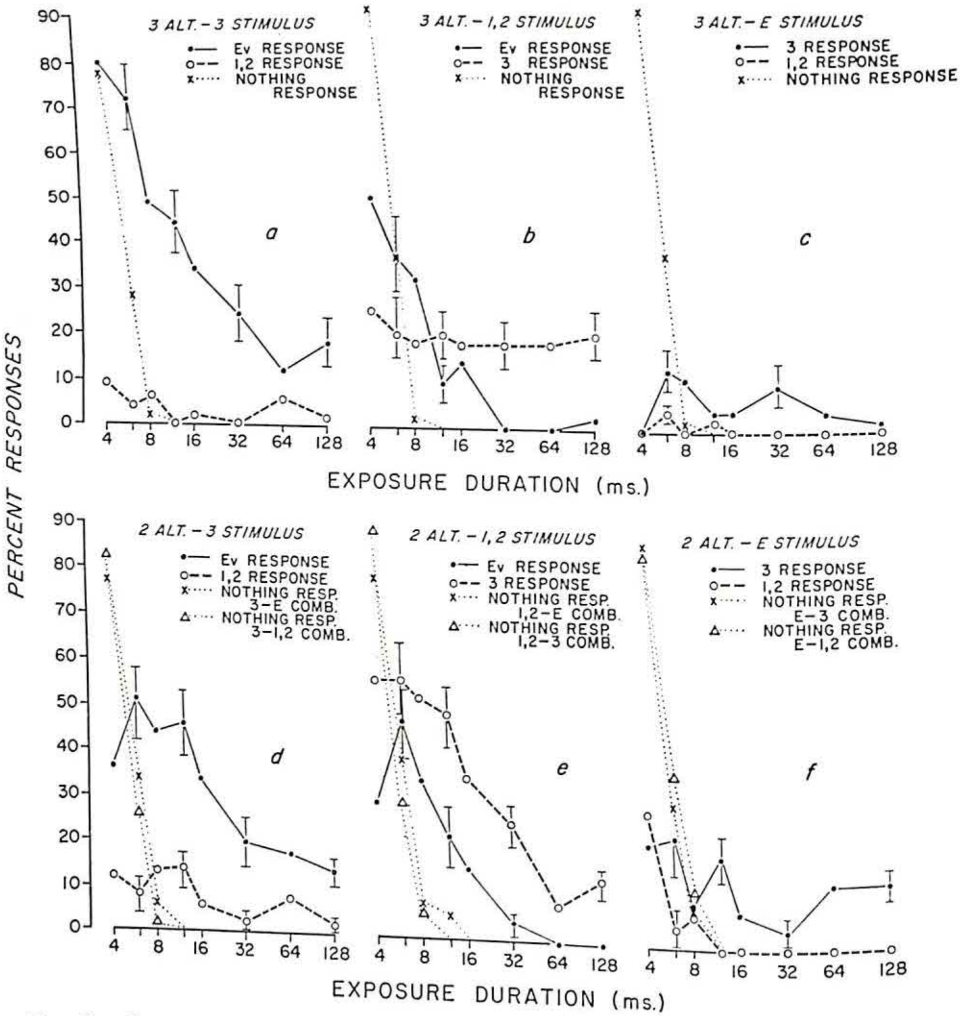


FIG. 2. Percent error responses to each stimulus as a function of exposure duration (solid and dashed curves). (Dotted curves show the percentage of "nothing" responses. Vertical bars show the *SE* of the percentage [see text]. Figures 2a, b, and c show responses to each pattern in the three-alternative condition. Figures 2d, e, and f show responses to each pattern in the two-alternative combinations.)

Representative values of the *SE* of the percentage are indicated by vertical lines at every second point of each error response curve in Fig. 2. These estimates were computed as $\sqrt{\frac{PQ}{N}}$, where P is the percentage of responses shown for a given point in Fig. 2, $Q = 1 - P$, and $N =$ the total number of responses excluding "nothing" responses, given to each pattern at the exposure duration indicated by the point.

Three-Alternative Condition

Ev pattern.—Since there are no variations in spacing within the Ev pattern, variation in exposure duration should not, according to Assumption *a* (see Introduction), affect its discrimination, and ideally all responses to it should be Ev responses.

As shown in Fig. 2c, there were indeed very few erroneous responses to this pattern.

3 pattern.—According to Assumption *b*, decreasing exposure durations will decrease the probability that the difference between spaces will be discriminated. Given all three patterns as possible alternatives, we would therefore expect a high incidence of Ev responses to the 3 pattern at the briefest exposures. Since decreasing exposure duration is presumed only to decrease the probability of perceiving existing differences between spaces, the 1,2 response should not occur to this pattern. Figure 2a confirms these predictions in showing that error responses to this pattern were limited largely to Ev responses. The upper half of Fig. 3 shows responses to this pattern for individual *S*s. These functions are clearly consistent with Fig. 2a.

1,2 pattern.—As for the 3 pattern, differences between spaces should not be perceived at the briefest exposures, according to Assumption *b*. Under these conditions, we expect Ev responses to be the predominant error. Figure 2b (solid curve) shows this to be the case. Frequency of Ev responses decreased sharply, as exposure duration lengthened.

Assumptions *b* and *c* imply that large differences between spaces will be the first to be perceived as a function of increasing exposure duration. At exposures of intermediate length, therefore, only the larger difference between spaces (4 vs. 8 mm.) in the 1,2 pattern is likely to be perceived, and it seems reasonable to suppose that this perception more closely approximates the 3 pattern than the other alternatives. In any case, Ev is not a probable response under these conditions, because spacing differences should be perceived, so that the 3 response should be the predominant error response at exposures of intermediate length. Accordingly, the 3-response function to the 1,2 pattern should look like an inverted V, first increasing, then decreasing with lengthening exposure duration. The 3-response curve in Fig. 2b (dashed line) is quite flat; however, this result is attributable to considerable individual differences in the exposure duration required for the emergence of the 3-pattern response. The lower half of Fig. 3 shows the responses to the 1,2 patterns in the three-alternative condition for each of the five *S*s. (Subjects 2,3,5 did not detect the 1,2 stimulus at 4 msec. This is indicated in Fig. 3 by breaking the initial portion of the solid curves for these *S*s.) For all *S*s the percentage of Ev responses (solid lines) drops off fairly quickly with lengthening exposures. As expected, the modal error percentage for the 3 response (dashed lines) occurs, in each case, at a longer exposure than the mode for the Ev response. Thus, the error curves for individual *S*s are consistent with the assumptions.

Two-Alternative Combinations

The results are shown in Fig. 2d, 2e, and 2f. These figures were drawn so as to be directly comparable to the three-alternative results. Note that the error curves within Fig. 2d, 2e, and 2f are derived from different two-alternative combinations. For example, the 3 pattern was a stimulus in the 3-Ev as well as the 3-1,2 combinations. Accordingly, the Ev and 1,2 error response curves to the 3 pattern shown in Fig. 2d derive from responses given in these different combinations. Also, since the 3 pattern occurred in two combinations, two distributions of "nothing" responses were obtained, as shown in Fig. 2d. Responses to the 1,2 and Ev

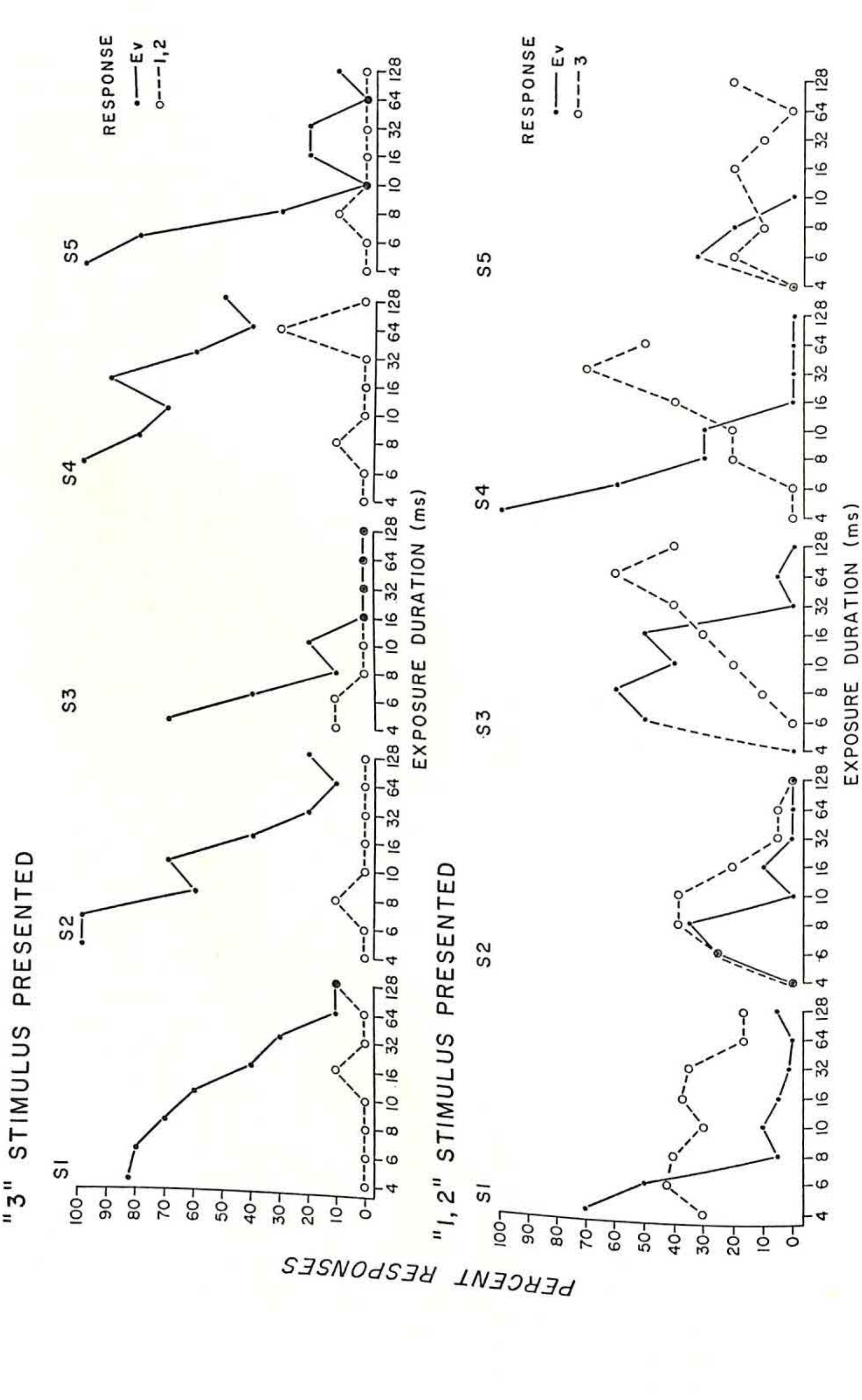


FIG. 3. Percent error responses as a function of exposure duration for each S when the 3 pattern (top half of figure) or 1,2 pattern (bottom half) was shown in the three-alternative condition.

patterns are shown in Fig. 2e and 2f, respectively.

Assumptions for the three-alternative condition led to the prediction of a 3-point, temporally determined ordinal scale of probable perceptions for the 1,2 pattern. That is, the probability of perceiving this pattern as evenly spaced was considered maximal at the briefest exposures, leading to a high proportion of Ev responses. At exposures of intermediate length, only the largest difference between spaces was likely to be perceived, leading to 3 responses as the most likely error. At longer exposures, when all differences between spaces are likely to be perceived, 1,2 responses should be maximal. Perception of the 3 pattern can be described by the first 2 points on the above scale. Thus, what is perceived in the three-alternative condition presumably approximates the available alternatives, but the perceptions as such should be independent of available responses. In fact, perceptions of the 1,2 and 3 patterns at some durations in the two-alternative combinations should *not* approximate the spacing characteristics of available alternatives. For example, both the 1,2 and the 3 patterns should often appear as evenly spaced at brief exposures, but the Ev pattern is not an alternative in the 1,2-3 combination. We propose a simple decision rule that will operate under these conditions: *S* will choose the available alternative which is closest to what he perceives along the temporally determined perception scale specified above.

1,2 pattern.—In the 1,2-3 combination, perceptions of the 1,2 pattern approximating the Ev pattern will lead to 3 responses rather than 1,2 responses, according to the decision rule. In this combination, 3 responses to the 1,2 pattern thus come from two presumed percepts. One is from the Ev perception just discussed. The other source of 3 responses derives from responses to exposures of intermediate length (see above). The proportion of 3 responses to the 1,2 stimulus in the 1,2-3 combination should, therefore, equal the sum of 3 and Ev responses to the 1,2 stimulus obtained in the three-alternative condition.

The dashed curve in Fig. 2e shows, in fact, a higher percentage of 3 responses to the 1,2 stimulus in the 1,2-3 combination than in the three-alternative condition (dashed curve, Fig. 2b). The assumed addition of Ev and 3 responses will be discussed further below.

In the 1,2-Ev combination the 3 response is not an available alternative at intermediate exposures, when only the largest spacing difference is presumed to be perceived. Under these conditions the 1,2 response should be selected because it is the only alternative containing some spacing differences. Accordingly, the 1,2 pattern should be correctly identified, except when it is perceived as equally spaced. The proportion of Ev responses should therefore be the same as that obtained in the three-alternative condition. Comparison of Fig. 2e and 2b (solid lines) shows, in fact, that the Ev response functions to the 1,2 stimulus were very similar in the 1,2-Ev and in the three-alternative condition. In both cases Ev responses occurred largely at brief exposures.

3 pattern.—Assumption *b* leads to the same response prediction for this pattern in the 3-Ev combination as in the three-alternative combination since, as before, this pattern is perceived either correctly or as equally spaced, depending on exposure duration. The solid lines in Fig. 2a and 2d show that the two functions are indeed very similar, except at the briefest exposures, where there were

TABLE 1

Obtained and Predicted Error Response Proportions to Each Stimulus in Two- and Three-Alternative Combinations

S-R	Three-Alt. Obt. Prop.	Two-Alt. Obt. Prop.	Two-Alt. Prediction	
			Ours	CRR
3-Ev	.426	.376	.426	.439
Ev-3	.057	.112	.057	.050
Ev-1,2	.006	.019	.006	.006
1,2-Ev	.180	.230	.180	.222
3-1,2	.027	.084	.027	.046
1,2-3	.188	.409	.368	.229

more Ev responses in the three-alternative condition.

In the 3-1,2 combination, perception of the 3 stimulus as evenly spaced should lead to 3 responses according to the above decision rule. As in the three-alternative condition, no 1,2 responses to this pattern should occur, so that in this combination no incorrect responses should occur to the 3 pattern. Figure 2d (dashed curve) shows that few error responses were, in fact, obtained.

Ev pattern.—As in the three-alternative condition, no errors should occur to the Ev pattern in the two-alternative combinations. Figure 2f shows somewhat more error responses to the Ev stimulus in the 3-Ev and 1,2-Ev combinations than were obtained in the three-alternative condition, suggesting a slight response bias. Errors to the Ev pattern (Fig. 2f) are still negligible relative to errors to the 3 and 1,2 patterns (Fig. 2d and 2e).

Table 1 shows another way of comparing the results of the two- and three-alternative conditions. The first column of this table identifies each possible error response to each pattern in either the two- or three-alternative conditions. The second column shows the obtained proportion, pooled over the five lowest exposure durations, of each possible error response to each stimulus pattern in the three-alternative condition. The third column contains similarly obtained error response proportions to each pattern in each two-alternative combination. The fourth column lists predictions of response probabilities in the two-alternative combinations made on the basis of the above discussion. To summarize briefly, we predicted that error response proportions in the two-alternative combinations should be identical with those in the three-alternative condition except for the 3 responses to the 1,2 stimulus. The proportion of 3 responses given the 1,2 stimulus in the 1,2-3 combination should be the sum of Ev and 3 responses (.180 + .188, see Column 2 of Table 1) given the 1,2 stimulus in the three-alternative condition. Results shown in Table 1 indicate that all of our predictions come within 5% of the obtained results.

Table 1 is also of interest because our predictions from the three- to the two-alternative conditions can be compared to those which follow from the Constant Ratio rule (CRR). This rule was first developed by Clarke (1957) and can be derived from Luce's (1963) choice theory. A number of studies (e.g., Hodge & Pollack, 1962) have demonstrated good predictive power for this rule. The CRR asserts that the ratio of any two response probabilities to a given stimulus is a constant regardless of the number of possible alternative responses to the stimulus. The CRR predictions must be obtained empirically from the responses to a master set of stimulus alternatives, and the rule is not concerned with the stimulus determinants of these responses. Thus, the obtained responses shown in Table 1 for the three-alternative condition were used to predict responses to each pattern in every two-alterna-

tive combination.[2] Column 5 of Table 1 shows these CRR predictions. On the basis of logical and empirical considerations the CRR predictions should be as accurate as ours except for prediction of the percentage of 3 responses to the 1,2 pattern in the 1,2-3 combination.

Here, the obtained results closely approximated our prediction that the 3 responses should correspond to the sum of Ev and 3 responses obtained to the 1,2 pattern in the three-alternative combination. The CRR prediction in this case was off by about 18%.

Discussion

The results of the present study indicate that use of rather simple assumptions about spatial and temporal characteristics which determine the perception of stimuli can effectively predict choice behavior. This formulation can predict response probabilities to each stimulus from master to subset (three to two-alternative combinations) as well or better (in one instance) than the CRR rule. Further, it is difficult to see how existing formulations of choice behavior could predict the directional, nonsymmetric order of responses which we predicted and obtained (e.g., more 3 responses to the 1,2 pattern than 1,2 responses to the 3 pattern). To be sure, "response tendencies" could be used as separate post-hoc descriptions of each asymmetric response tendency obtained, but such notions are hardly useful for the a priori specification of these stimulus-response relations.

One advantage of the use of these stimulus-derived as compared to empirically obtained response predictions is that the former postulate more specific behavioral determinants of choice behavior. Attention to such stimulus variables in decision theories might therefore not only increase the generality of their admittedly powerful formulations, but help clarify the behavioral bases of the decision process itself.

[2] The predictions from Clarke's rule were obtained by using a derivation from the rule given by Anderson (1959). The derived formula reads as follows,

$$P(r_i/s)_S = \frac{P(r_i/s)_M}{\sum_{r' \epsilon S} P(\bar{r}'/s)_M}$$

That is, $P(r_i/s)_S$, the probability of response r_i to stimulus s_i in the subset, S, is equal to $P(r_i/s)$ in the master set, M, divided by the sum of the probabilities obtained in M for all responses to s which were possible in S. As applied to our data, the predicted CRR proportion of 3 responses to the 1,2 stimulus would be

$$P(3/1,2)_{\substack{\text{Two}\\ \text{Alternative}}} = \left[\frac{P(3/1,2)}{P(3/1,2) + P(1,2/1,2)}\right]_{\substack{\text{Three}\\ \text{Alternative}}}$$

whereas our prediction would be,

$$P(3/1,2)_{\substack{\text{Two}\\ \text{Alternative}}} = [P(3/1,2) + P(E/1,2)]_{\substack{\text{Three}\\ \text{Alternative}}}.$$

References

Anderson, C. D. The constant ratio rule as a predictor of confusions among visual stimuli of brief exposure duration. Technical Note Document No. AD-160706, Armed Services Technical Information Agency, May 15, 1959. (Mimeo)

Attneave, F. *Applications of information theory to psychology.* New York: Holt, 1959.

Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. Acoust. Soc. Amer.*, 1957, **29**, 715–720.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Hodge, M. H., & Pollack, I. Confusion matrix analysis of single and multidimensional auditory displays. *J. exp. Psychol.*, 1962, **63**, 129–142.

Kaswan, J. Tachistoscopic exposure time and spatial proximity in the organization of visual perception. *Brit. J. Psychol.*, 1958, **49**, 131–138.

Kaswan, J., & Young, S. Stimulus exposure time, brightness, and spatial factors as determinants of visual perception. *J. exp. Psychol.*, 1963, **65**, 113–123.

Luce, R. D. Detection and recognition. In R. D. Luce, E. Galanter, & R. R. Bush (Eds.), *Handbook of mathematical psychology.* Vol. 1. New York: Wiley, 1963.

(Received March 6, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 450–459

DISTRIBUTION AND SEQUENCE EFFECTS IN JUDGMENT [1]

ALLEN PARDUCCI AND ARTHUR SANDUSKY

University of California, Los Angeles

In a visual-recognition task, the proportion of trials for which a light flashed in the left of 2 closely spaced positions was either .2, .5, or .8. The proportion of correct identifications of each alternative varied inversely with its proportion of presentation. Most of this variation could be deduced from the 1st-order stimulus dependencies, Ss being more accurate following stimulus alternations than following stimulus repetitions. This suggests that the successive stimulus pairs were the effective stimuli. The results appear inconsistent with the data from studies of signal detection. Except when informational feedback is given, the sequential effects also appear inconsistent with a learning model for signal detection. A simple Thurstone model, elaborated to include a criterion band, describes both the stimulus and the response dependencies.

Judgment implies comparison. "Good" and "bad," "large" and "small" compare what is being judged with some contextual standard. The child is told that he is "bad" when his behavior is worse than it has been in the past or worse than the behavior of other children. For such value judgments, the particular context that provides the standard may be very complex. For simple perceptual phenomena, however, various contextual theories of judgment account for the major effects of the relevant physical context (e.g., Helson, 1959; Johnson, 1955; Parducci, 1963).

These contextual theories predict the judgments from the *stimulus distribution*, the relative frequencies with which the different physical values are presented. For example, a stimulus at the midpoint of the range is judged "large" if the smaller stimuli are presented more frequently, "small" if the larger values are presented more frequently. But when presentation is successive, one stimulus at a time, the overall distribution may not have the same effects upon the judgment of an early stimulus as on a later one. The end points of the scale of judgment cannot be established until the stimulus extremes have been presented, and the effects of these extreme values may diminish as a result of subsequent presentations. More generally, the judgments of any stimulus may depend upon the particular stimuli that precede it.

Contextual theories have predicted the mean judgments averaged over the entire series, ignoring possible trial-to-trial changes. However, these changes may also reflect the effects of the stimulus distribution. Sequential changes may even explain overall distribution effects.

The evidence for trial-to-trial sequential effects is sparse for category judgments (e.g., Garner, 1953; Parducci & Marshall, 1962). Typically, the stimulus extremes are presented early in the series, stabilizing the judgments so that later shifts are relatively small. Since there are usually at least five different stimuli, the different subsequences may not

[1] This research was supported by a grant (M-5465) from the National Institute of Mental Health, United States Public Health Service. Special thanks are due Norman H. Anderson for a critical reading of the manuscript.

occur with sufficient frequency to permit assessment of sequential effects.

Preliminary research developed a simplified task in which the scale of judgment is never stabilized (Parducci, 1964). With just two stimuli, enough data could be obtained to permit more adequate analysis of the effects of different subsequences of stimuli. Lifted weight, visual brightness, and both loudness and pitch yielded similar results: judgments appeared to be largely determined by the two most recent presentations, and *S*s were more accurate on stimulus alternations than on stimulus repetitions. The present report is of a more extensive study using judgments of visual position.

METHOD

Stimulus display.—The experimental room was completely dark, except for stimulus lights displayed on a vertical panel, 5 ft. in front of *S*. The panel had three $\frac{1}{8}$-in. holes, each centered in front of a $\frac{1}{4}$-w. neon bulb, Type 45, in series with 20,000-ohm resistance. Two of the lights were $\frac{1}{2}$ in. apart on a horizontal line, approximately at the eye level of the seated *S*. The third was 11 in. higher and centered to complete an isosceles triangle with the other two, forming an angle of approximately 3° at the apex.

Programing and recording.—The upper light was always on, providing a constant reference point for *S*. The two lower lights were activated in predetermined sequences, with one or the other flashing every third second for 185 msec. Sequences of 300 presentations were controlled by punched tape, the tape reader closing the alternative circuits at intervals set up by motor-driven cams. The relative position of each flash was indicated by *S* who pushed a toggle switch either to the left or the right. Both tape reader and toggle switch were connected to an event recorder which provided a record of the successive stimuli and the associated responses.

Instructions.—After being seated in the experimental room, *S* was read the following instructions:

I am measuring your ability to discriminate between lights which appear in two different lateral positions. The room will be darkened, and you will see this fixed dot of light which will remain on throughout the experiment. Once every 3 seconds, a second light will flash briefly below this dot, in one of two fixed positions. Your task is to judge in which of the two positions the lower flash appears. If the flash were in the left position, you would push this lever to the left; if in the right position, to the right. The purpose of the constant light is to help you keep your general spatial orientation in the dark; it does not necessarily bisect the positions you are judging; rather, it appears in a different relationship to these two positions for different people. Head movements, for example, are likely to change the apparent relationship between the top light and the two flash positions. Your performance will be most accurate if you try to fix your attention on the general region where the flashes will occur, using the top light to keep your eyes from wandering between flashes. Always respond to each flash of light before the next flash occurs—even if you have to guess, which, of course, you will have to do for the very first presentation. Although it is a fine discrimination, you will be able to do better than chance. The whole series will take 15 minutes.

Conditions.—Three distributions of stimuli were used, the left light being presented with probability .2, .5, or .8. Two randomized orders of presentation were used for each schedule, and the left presentations were balanced in each of the two blocks of 150 presentations. The number of alternations between left and right stimuli were also fixed at the expected value for each block, 75 for the .5 and 30 for the .2 and .8 distributions. The number of homogeneous stimulus runs of different length corresponded closely to the expected values for each of the blocks.

An additional condition, .5F, employed the .5 distribution with informational feedback. The correct position, left or right, was announced after each incorrect response, *E* making the correction over an intercom from the adjoining room.

Subjects.—The *S*s were 160 undergraduate volunteers, 40 being assigned to each distribution and 40 to the .5F condition. Assignment was evenly divided between the orders of presentation for each condition, 10 of the 20 *S*s exposed to each order being males. Additional data will be reported for a single, highly practiced *S*.

Statistical analyses.—For Fig. 1 and 2 and

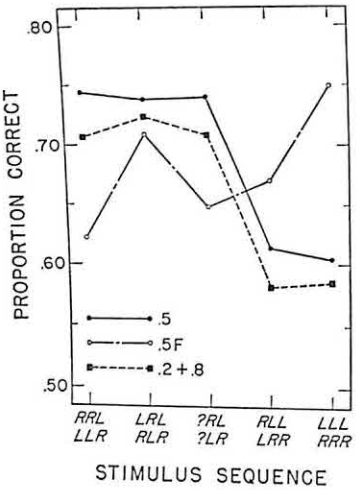


FIG. 1. Mean proportion of correct responses following different subsequences of stimuli, averaged for complementary subsequences.

also for several of the statistical analyses, unweighted means were taken when averaging proportions for complementary stimulus sequences (such as left, left, right; and right, right, left) so as not to reflect *Ss*' position preferences. The major analyses of variance included Distribution, Order of Presentation, and Sex as independent variables. Neither Order of Presentation nor Sex had significant effects, except that males were consistently more accurate; and the data are not reported separately for these two variables. All significant comparisons ($p < .05$) are reported.

RESULTS

The major results of this experiment reduce to a single rule: accuracy is greater on trials for which the position of the flash has shifted. It is easier for *S* to discriminate the direction of the shift than to keep his bearings when the flash recurs in the same position. This greater accuracy for alternations than for repetitions is consistent with common sense, with the results of the preliminary studies using other stimulus materials, and with various models for judgment (Parducci, 1964). What is unexpected is the degree to which the many complex features of the data reduce to this simple difference. Unless informational feedback is provided, *S*'s response is almost completely determined by the physical relationship between the presented stimulus and the immediately preceding stimulus. Contrary to the stimulus pooling postulated by contextual theories of judgment, *S*'s response appears to be independent of even the second stimulus back. And yet, as will be shown in the sections which follow, the usual effects of stimulus distribution are obtained.

Stimulus Dependencies

The major experimental finding is illustrated in Fig. 1. Each point on the three curves represents the proportion of correct responses, $P(c)$, as a function of sequences of the three immediately preceding stimuli (including the stimulus to which the response is given). Thus the leftmost point on each curve represents accuracy for an alternation trial preceded by one or more repetition trials (*RRL* indicating a sequence for which the order is right, right, left; its complement is *LLR*).

Consider first the curve for the .5 condition. The two leftmost points, and also the third which is their weighted mean, have approximately equal values. Alternation trials are all the same: accuracy is unaffected by the length of the homogeneous stimulus run preceding the alternation.

Accuracy on repetition trials is shown by the two rightmost points on the same curve. Although accuracy is much lower for repetition trials, it is the same for all repetition trials, i.e., for homogeneous runs of either two or else three or more stimuli. The response on a given trial is again independent of the second stimulus back.

TABLE 1

MEANS AND *SD*s OF RESPONSE PROPORTIONS

	.2		.8		.5		.5F	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
$P(l)$	.48	.10	.53	.13	.52	.09	.52	.08
$P(c)$	.61	.09	.61	.09	.67	.07	.68	.08
$P(c\mid L)$	.73	.13	.59	.13	.70	.10	.70	.12
$P(c\mid R)$	.59	.11	.69	.15	.65	.11	.66	.10
$P(c\mid LL)$	.66	.18	.56	.14	.63	.13	.74	.12
$P(c\mid LR)$	.66	.16	.72	.15	.71	.12	.63	.14
$P(c\mid RL)$	.75	.14	.70	.13	.77	.10	.66	.14
$P(c\mid RR)$	.56	.11	.55	.21	.59	.14	.68	.10

The same inferences can be made for the .2 and .8 conditions. Their combined curve, representing unweighted means of their respective proportions, closely resembles the .5 curve. Although the .2 + .8 curve is uniformly lower, the differences between conditions are small relative to the effects of sequences within conditions. Sequences of more than three stimuli occur with insufficient frequency to permit reliable assessment of sequential effects. However, no clear trends were noted with the longer series. For instance, $P(c)$ for the combined .2 and .8 conditions does not drop below .55 for homogeneous runs of up to length 5.

The .5F curve suggests quite different inferences, the general effect of informational feedback being to increase the probability of the response that would have been correct on the preceding trial. Treatment of the data from the .5F condition will be reserved for the end of the Results section. The immediate objective here is to demonstrate the statistical significance of the major results for the conditions in which feedback was not given.

In the absence of informational feedback, accuracy on Trial n depends upon the immediate stimulus pair, S_{n-1} and S_n. The means and *SD*s of these four conditional proportions, $P(c\mid S_nS_{n-1})$, are presented in the bottom rows of Table 1. For example, the last row, labeled $P(c\mid RR)$, gives the proportion of correct responses following repetitions of R. These four values were combined for each S to provide the raw scores for several analyses of variance.

The first analysis assessed the difference between alternation and repetition trials. Each S's score was the sum of $P(c\mid LR)$ and $P(c\mid RL)$, minus the sum of $P(c\mid LL)$ and $P(c\mid RR)$. The mean difference, .126, is highly significant, $F\ (1, 114) = 141.38$. Of the 120 *S*s, 107 were more accurate on alternation than on repetition trials.

The second analysis tested position preference. The difference in accuracy for the two stimuli, L and R, was indicated by the sum of $P(c\mid RL)$ and $P(c\mid LL)$, minus the sum of $P(c\mid LR)$ and $P(c\mid RR)$. The mean for this measure of position preference, .045, differs significantly from zero, $F\ (1, 114) = 5.85$, indicating that *S*s were more accurate on L. A third analysis, based on the sum of $P(c\mid LL)$ and $P(c\mid LR)$, minus the sum of $P(c\mid RR)$ and $P(c\mid RL)$, yielded no significant effects. There is thus no evidence of interaction between the position of the presented stimulus, L or R, and the stimulus relationship, alternation or repetition.

Distribution effects.—Thus far, the only indication of distribution effects has been the difference in Fig. 1 between the .2 + .8 and the .5 curves. Using the sum of each S's four conditional proportions as a measure of overall accuracy, small but significant differences were found for both Distribution, $F\ (2, 114) = 3.13$, and Sex,

F (1, 114) = 5.33, the better performance coming from males and with the .5 condition. Of course, P (c) is greatest for the .5 condition because stimulus alternations are more frequent than in the .2 and .8 conditions.

The overall proportion of left responses, $P(l)$, provides another measure of distribution effects. Alternative models for the theory of adaptation level agree in predicting that $P(l)$ would be independent of Distribution (Parducci, 1964). As shown in Table 1, this contrary-to-common-sense prediction is very nearly confirmed. The variation in $P(l)$ is only $\frac{1}{12}$ the variation in $P(L)$. However, the variation in $P(l)$ is statistically significant, F (2, 108) = 3.23.

The differences on these overall measures seem minor. The major effects of stimulus distributions are found when judgments are tabulated separately for each stimulus in accordance with the traditional procedure for assessing context effects in judgment. As shown in Table 1, the mean proportion of correct responses to L, $P(c|L)$, varies inversely with $P(L)$, F (2, 117) = 14.34; and $P(c|R)$ varies directly with $P(R)$, F (2, 117) = 7.14. The directions of these variations correspond to those obtained with category judgments when the frequency distribution of the stimuli is varied (Parducci, 1963). In the present case, accuracy for each stimulus varies directly with the proportion of trials on which it was preceded by the other stimulus. Thus, the major distribution effects follow directly from the proportions of correct responses to the different stimulus pairs.

Stationarity.—There was some loss of accuracy over the series under all conditions. Although the mean drop in $P(c)$ between the two halves was only .03 and the mean drop in $P(l)$ only .02, individual *S*s shifted dramatically. For one S, $P(l)$ dropped from .88 for the first half to .21 for the second half. The mean absolute shift in $P(l)$, irrespective of direction, was about four times as great as the mean algebraic shift for the various conditions.

Chi squares were computed for each S to assess the stationarity of his first-order transition matrix (Suppes & Atkinson, 1960, p. 56). Separate tabulation was made for the two halves of the sequence. Each of eight chi squares tested the significance of one of the shifts in response frequencies between halves. Each test applied to one of the two stimuli, conditional on one of the four preceding stimulus-response combinations. These eight values were summed for each S. More than two-thirds of the *S*s had significant overall chi-square values, indicating a general departure from stationarity. Inspection of the transition matrices suggests that different *S*s shift in opposite directions, canceling each other out in the various group analyses. The data of this section suggest that the individual deviations from stationarity are mainly shifts in response preferences which are not reflected in the systematic mean shifts for different conditions.

Response Dependencies

Responses tend to occur in homogeneous runs. Response dependency is indicated by the proportion of responses that repeat the immediately preceding response. The means of the proportions of these first-order repetitions are .61, .58, and .64 for the .2, .5, and .8 conditions. These means are greater than would be expected from the stimulus dependencies, reported in the previous section for the two stimuli and four stimulus pairs. The expected proportion of response repetitions is given by the following equation:

$$\begin{aligned} P(J_n = J_{n-1}) = P(LL)[P(l|L)P(l|LL) + P(r|L)P(r|LL)] \\ + P(LR)[P(l|L)P(l|LR) + P(r|L)P(r|LR)] \\ + P(RL)[P(l|R)P(l|RL) + P(r|R)P(r|RL)] \\ + P(RR)[P(l|R)P(l|RR) + P(r|R)P(r|RR)]. \quad [1] \end{aligned}$$

Substitution of the response proportions from Table 1 into Equation 1 yields the expected proportions, .49, .48, and .49. With 90% of *S*s exceeding the overall expected value for

their condition, it is clear that the preceding response is repeated much more frequently than would be predicted from the overall stimulus dependencies.

The means of the proportions of second-order response repetitions, i.e., of trials for which the response repeats the second response back, are .62, .59, and .66. The mean difference favors the second-order dependency for all three conditions, $F(1, 117) = 4.15$. The expected value for second-order repetitions, computed using an analog of Equation 1, is close to .50 for each of the three conditions.

It is surprising that *S*'s response should be more dependent upon his second response back than upon his immediately preceding response. However, the differences between the first- and second-order dependencies approximate the differences predicted from the stimulus dependencies. It can be shown that the predicted values are higher for the second-order dependencies whenever: $P(c) > .5$ for each stimulus and each stimulus sequence, and $P(c)$ is greater for stimulus alternations than for stimulus repetitions. These conditions hold for the group data and also for the great majority of individual *S*s in this experiment. Another factor is that any occasional tendency to alternate responses may decrease first-order repetitions but increase second-order repetitions.

Much of the difference between the predicted and obtained repetition proportions may depend upon individual differences in response preference. However, inspection of the individual scores indicates that marked response dependencies may be found even when there are no overall position preferences. For example, one *S* in the .5 condition had both first- and second-order repetition rates of approximately .70, even though his overall $P(l)$ and the values expected from his stimulus dependencies were each close to .50.

Analysis was also made of each *S*'s response dependencies on stimulus runs of length 3 or more (*LLL* or *RRR*). Even for those runs on which either one or the other of the preceding responses had been incorrect, the majority of *S*s were more often correct than wrong on Trial *n*. This indicates that the continuing accuracy on runs cannot be reduced to *S*'s tendency to repeat his previous responses.

Practice Subject

The major experimental findings were replicated using a single, highly practiced *S*. One undergraduate female was hired to serve

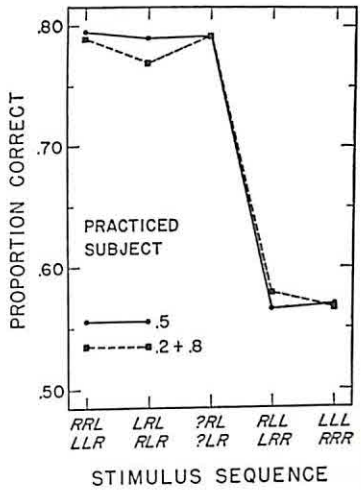


FIG. 2. Mean proportion of correct responses by practiced *S* following different subsequences of stimuli, averaged for complementary subsequences.

for 62 separate 300-trial sessions. Two to four sessions were given each day, spread over a month, with at least 15 min. between sessions. The three stimulus conditions were given in randomized order; the .5 condition was used for about half the sessions, and the other half was evenly divided between the .2 and .8 schedules. Informational feedback was never given.

Her proportion of left responses dropped systematically over the first half of the month before stabilizing on a moderate preference for the right. This drop was associated with decreased accuracy on .8 and increased accuracy on .2. There was no systematic drift in accuracy for the .5 condition, the value of $P(c)$ always falling between .60 and .72, with an overall mean of .673. This was virtually identical to $P(c)$ for the single-session .5 group. There was no clear evidence of interschedule sequential effects, e.g., greater $P(l)$ on .5 following .2 than following .8. Nor was there any systematic drift in her proportion of first-order response repetitions (calculated only for the .5 schedule). Although the overall proportion was only .507, this was significantly above the expected value (based on her overall stimulus dependencies).

As shown in Fig. 2, the stimulus dependencies for the practiced *S* correspond closely to the dependencies of single-session *S*s. As with Fig. 1, Fig. 2 can be summarized by just two values: (*a*) proportion correct on alternation trials, and (*b*) proportion correct on

repetition trials. However, when Fig. 1 and 2 are compared, it is seen that the practiced *S* shows a greater difference between alternation and repetition trials and does not show the lower accuracy for the .2 and .8 conditions.

Chi-square tests were performed upon the practiced *S*'s $P(c)$ in the .2 and .8 conditions for each of the four stimulus pairs, using the four proportions for the .5 schedule to obtain expected values. The shifts in proportions were statistically significant, χ^2 (8) = 28.32, $p < .001$.

As with single-session *S*s, the practiced *S* was less likely to be correct on the more frequent stimulus: her $P(c|L)$ values were .729, .635, .542 for the .2, .5, and .8 conditions. This replication of the important distribution effects is, of course, consistent with the dramatic difference in accuracy between alternation and repetition trials. Considered in conjunction with the stability in overall accuracy through the month, these results from a highly practiced *S* suggest that the major findings of this experiment do not depend upon some initial set brought into the session by naive *S*s.

Feedback

The data from the .5F condition were analyzed in the manner reported for the other conditions. As shown in Table 1, there was little difference between the .5F and .5 values for either $P(l)$ or $P(c)$. Indeed, despite the informational feedback, there was actually a mean drop of .023 in $P(c)$ between the halves of the sequence for the .5F condition, F (1, 39) = 5.48. Moreover, the .5F chi squares for stationarity closely resembled those for the .5 condition.

The failure of reinforcement to raise $P(c)$ may be clarified by comparison of the curves for the .5 and .5F conditions in Fig. 1. With feedback, the response is sharply affected by the second stimulus back. Accuracy increases on repetition trials with increase in the length of the homogeneous stimulus run, i.e., with additional reinforcements for the response that is now correct. Accuracy on stimulus alternations is much lower when the alternation is preceded by a stimulus repetition, i.e., by two or more reinforcements for the response which is now incorrect. The net result of the two reinforcement effects is to reverse the direction of the overall difference in accuracy between alternation and repetition trials, and on this comparison the .5 and .5F conditions were significantly different, F (1, 76) = 59.37.

The mean proportions of the first- and second-order response repetitions for .5F are .553 and .547, respectively. Both are lower than the corresponding values for the .5 condition, even though the expected values, .510 and .517, are higher. These changes are consistent with the other effects of feedback since the proportions for response repetitions should approach the proportion of stimulus repetitions, in this case .5, insofar as *S*s follow the preceding reinforcement.

Discussion

Adaptation-Level Theory

Definition of the stimulus.—Without informational feedback, the immediate stimulus pair is the effective stimulus for judgment. The many possible stimulus dependencies reduce to the proportions of correct responses for the four pairs: *LL*, *LR*, *RL*, and *RR*. Prediction of the response is not improved by taking into account even the second stimulus back. The pairs operate as four discrete stimuli. If the results are analyzed with respect to these four rather than to just two stimulus values, all sequential effects disappear.

This redefinition of the stimulus permits interpretation of the data in terms of frames of reference or adaptation levels which would be impossible with just two stimulus values, *L* and *R*. The difficulty would lie in the continuing difference between alternation and repetition trials, i.e., first-order stimulus dependencies, in spite of the complete

absence of even second-order dependencies. This would require extreme weighting for recency, and the frame of reference or adaptation level would thus be completely determined by the presented stimulus and the stimulus which immediately preceded it. But extreme weighting for recency would leave unexplained the differences in response to each stimulus pair for different stimulus distributions.

As in previous work (Engel & Parducci, 1961), theoretical simplification may be obtained by employing a more generous definition of the stimulus. The present proposal is to include the value of the immediately preceding stimulus and also a constant background value in the definition of the effective stimulus for judgment. The background value, B, is conceptualized as a reference point on the left-right dimension, based on cues provided by the reference light and interoceptive stimulation. B is introduced to account for *S*s' accuracy on repetition trials.

These physical values are combined so that $\dot{S}_n$, the *effective* stimulus on Trial n, is defined as the difference, S_n minus a weighted mean of B and S_{n-1}:

$$\dot{S}_n = S_n - [aB + (1 - a)S_{n-1}], \quad [2]$$

where a is an empirical constant representing the relative influence of the background and the preceding stimulus.

This specification of the effective stimulus for judgment is a variation on the typical equation for adaptation level (Helson, 1959). Here, $aB + (1 - a)S_{n-1}$ is analogous to the perceptual adaptation level, or neutral point on the left-right dimension. If $a = .5$, the successive differences between the effective values of *RL*, *LL*, *RR*, and *LR* must all be equal; the larger the value of a, the smaller the difference between the effective values of *RL* and *LL*, and also between *RR* and *LR*.

The analyses of variance reported for the stimulus dependencies indicated that the differences in correct response to the four effective stimuli can be broken down to two major sources: the greater accuracy on alternation trials, and the greater accuracy on L trials. The effects of these two sources may be considered additive since their interaction is not significant. Almost 85% of the variance between the 12 first-order contingencies, listed in the bottom four lines for the .2, .5, and .8 conditions in Table 1, can be accounted for by the following prediction equation:

$$P(c|S_n, S_{n-1}) = M + A + P, \quad [3]$$

where $P(c|S_n, S_{n-1})$ is the proportion of correct responses to each of the four effective stimuli. The values of the parameters, determined from the data in Table 1, were $M = .656$; $A = +.063$ for alternation trials, $-.063$ for repetition trials; and $P = +.023$ for L trials, $-.023$ for R trials.

The first constant, M, reflects the difficulty of discrimination and would be affected by environmental conditions such as the angle which L and R form with the reference light. The alternation constant, A, and the position-preference constant, P, are affected by the values of a and B in Equation 2. The redefinition of the stimulus thus permits this simple accounting for most of the variation in the stimulus dependencies for the .2, .5, and .8 conditions.

For the feedback condition, however, this simple account does not suffice. With feedback, there is increased accuracy on repetition trials but decreased accuracy on alternation trials. This seems reasonable since *S*s would be expected to follow the preceding reinforcement when they cannot discriminate differences between successive presentations. The proportions for the .5 and .5F conditions each differ by about .10. However, there are also second-order stimulus dependencies for the .5F condition. These would require additional theoretical analysis.

Distribution effects.—Two small distribution effects are found in accuracy for the stimulus pairs. First, the .2 + .8 curve in Fig. 1 is lower than the .5 curve. This difference is hard to explain, but it is not found for the practiced S in Fig. 2. Second, the difference between the proportions for the .2 and .8 conditions within each of the bottom four lines of Table 1 appears to be systematic. The *LL* and *RL* proportions are much lower for .8, but the *LR* value is lower for .2. Accuracy thus tends to be lower for pairs ending with the more frequently presented stimulus. Similar differences were found for the practiced S. This would seem to be a typical distribution effect, appearing as a consequence of S's tendency to equalize his use of the two responses (Parducci, 1963). Though much weaker than the major effects of distribution, these differences indicate that neither the present analysis, using stimulus pairs, nor

an adaptation-level analysis (Helson, 1959), using extreme weighting for recency, gives a complete account of the distribution effects.

The present analysis does account for the major effects of stimulus distribution. The conventional assessment of these effects would be made using the judgments of L and R, separately. Both $P(c|L)$ and $P(c|R)$ vary with stimulus distribution. The direction of the variation is consistent with the contextual theories of judgment which predict category shifts toward the more frequently presented stimulus. However, much of the present distribution effect follows directly from the first-order stimulus dependencies. Since $P(c|RL)$ is greater than $P(c|LL)$, $P(c|L)$ is necessarily greater for the .2 condition (which has four times as many RL as LL sequences) than for the .5 condition (where these two sequences have equal frequency) or for the .8 condition (which has four times as many LL as RL sequences).

Decision Theory

Signal detection.—Although it is not clear that the theory of signal detection applies to the present task (Tanner, 1956), the theory correctly describes the results of a similar experiment (Kinchla, 1964). Kinchla found that the proportion of correct responses for either a left or right position of a threshold level, visual stimulus varied *directly* with the proportion of stimulus presentations in that position. However, $P(c|L)$ varies *inversely* with $P(L)$ in the present experiment, and $P(c|R)$ varies inversely with $P(R)$. The sequential effects were also different for Kinchla's detection task. His data were fitted by a learning model in which $P(c)$ increases on stimulus repetitions, as in the present .5F condition (Atkinson, 1962). Neither the overall shifts in detection rate nor the sequential effects attributed to learning are found with the present, nondetection discrimination, unless there is feedback.

Perhaps the critical difference between the tasks is in the way the responses are defined. In the signal-detection task, the problem is to perceive the occurrence of a signal. In the present task, S always perceives the flash, his problem being to discriminate its location. This discrimination requires some kind of comparison, L being left with respect to R. The usual signal-detection task seems, at least formally, to be less dependent upon comparisons between alternative stimuli.

Band-criterion model.—The traditional Thurstone model of the stimuli, postulating overlapping discriminal distributions, can be elaborated to account for both the stimulus and the response dependencies in the present data. Assume that S establishes a band criterion such that discriminal values to the left and right of the band are responded to as "left" and "right," respectively. For each stimulus, let β denote the proportion of presentations with discriminal values outside the band. Assume for the present analysis that the distribution overlap is negligible in the regions outside the criterion band so that S is correct on all such observations.

When the discriminal value of the stimulus falls within the band, accuracy is assumed to depend upon the preceding stimulus and response. If the discriminal value of the preceding stimulus fell outside the criterion band, S would discriminate the relationship between the two discriminal processes, and reverse his preceding response. If the value of the preceding stimulus fell within the criterion band, S would repeat his preceding response with fixed probability, α. From the foregoing assumptions, it follows that $P(c)$ for stimulus repetitions with the .5 schedule is given by $\beta + \alpha(1-\beta)^2/(1+\beta)$ when the preceding response was correct, and by $1-\alpha(1-\beta)$ when the preceding response was incorrect. The corresponding $P(c)$ values for stimulus alternations are given by $1-\alpha(1-\beta)^2/(1+\beta)$ and $\beta+\alpha(1-\beta)$. Corresponding expressions for these values were derived for the .2 and .8 schedules.

Assuming that α and β are constant for all three schedules, their values were estimated to be .83 and .26, respectively, by a least-squares approximation to the 24 first-order proportions, $P(c|S_{n-1},$

J_{n-1}, S_n). The mean absolute deviation between fitted and obtained values is only .05, with 88% of the variance between the 24 proportions being accounted for using only two empirical constants.

The single parameter, β, accounts for most of the difference in accuracy between alternation and repetition trials and also for the overall level of accuracy. The response-repetition parameter, α, enables the model to account for the difference of .26 in $P(c)$ between trials on which the stimulus corresponds to, and trials on which it differs from the preceding response. The model also captures the additional distribution effects reflected in the relative drop in accuracy for stimulus pairs ending with the more frequently presented stimulus. The power of the band-criterion model lies in its ability to describe this difference and also the stimulus dependencies with just two empirical constants.

Whether the data from this experiment are described using the band-criterion model or with respect to stimulus pairs, most of the effects of stimulus distribution can be deduced from first-order sequential effects. Sequential data have been virtually ignored by the various theories of judgment. The success of this demonstration should encourage analysis of sequential phenomena in experiments with additional stimulus and response alternatives. It would be gratifying to discover that the major context effects in more complex and perhaps better anchored situations could also be reduced to the finer grain of the data.

REFERENCES

ATKINSON, R. C. Mathematical models in research on perception and learning. In M. Marx (Ed.), *Psychological theory.* (2nd ed.) New York: Macmillan, 1962.

ENGEL, G., & PARDUCCI, A. Value of background in the specification of the stimulus for judgment. *Amer. J. Psychol.*, 1961, **74,** 569–575.

GARNER, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46,** 373–380.

HELSON, H. Adaptation-level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

KINCHLA, R. A. A learning factor in visual discrimination. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford Univer. Press, 1964.

PARDUCCI, A. Range-frequency compromise in judgment. *Psychol. Monogr.*, 1963, **77**(2, Whole No. 565).

PARDUCCI, A. Sequential effects in judgment. *Psychol. Bull.*, 1964, **61,** 163–167.

PARDUCCI, A., & MARSHALL, L. M. Assimilation vs. contrast in the anchoring of perceptual judgments of weight. *J. exp. Psychol.*, 1962, **63,** 426–437.

SUPPES, P., & ATKINSON, R. C. *Markov learning models for multiperson interactions.* Stanford: Stanford Univer. Press, 1960.

TANNER, W. P., JR. A theory of recognition. *J. Acoust. Soc. Amer.*, 1956, **28,** 882–888.

(Received February 19, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 460–466

MEDIATION INSTRUCTIONS VERSUS UNLEARNING INSTRUCTIONS IN THE A-B, A-C PARADIGM [1]

KENT M. DALLETT AND LESTER D'ANDREA

University of California, Los Angeles

In 3 experiments, Ss learned the 1st list of an A-B, A-C paradigm, and were then differently instructed for List 2 learning. Mediation Ss were told to use the List 1 responses as aids to List 2 learning, while Unlearning Ss were told to unlearn List 1 responses in order that they not interfere with List 2 learning. The differential instructions did not affect the rate of List 2 learning, but did affect the relative difficulties of List 2 items. The Mediation Ss were slightly better in subsequent List 1 recall, and reported many more instances of interlist mediation than did the Unlearning Ss.

In an influential study of transfer and retroaction effects in the A-B, A-C, and A-B, A-B′ paradigms (in which, respectively, two successive lists have dissimilar—B, C—or similar—B, B′—responses, and identical stimuli), Barnes and Underwood (1959) proposed that *S*s commonly unlearn List 1 responses when they are dissimilar to List 2 responses. However, when List 2 responses turn out to be similar to List 1 responses, the List 1 responses are used as mediators of the List 2 responses, and are therefore preserved from unlearning. As one goes from dissimilarity to similarity of responses, unlearning presumably gives way gradually to mediation. Evidence has recently been provided that some unlearning can occur in what appears to be an A-B, A-B′ paradigm (Dallett, 1964): the experiments to be reported were designed to explore the role of mediation in the A-B, A-C paradigm. Specifically, an attempt was made to vary the relative predominance of mediation and unlearning by means of differential instructions.

The three experiments to be reported are essentially similar in plan. After learning the first (A-B) list for a fixed number of presentations, *S*s were divided and differently instructed for List 2 learning. Half of the *S*s (the Mediation or "M" condition) were told that they would do best to seek some connection, however bizarre, between the List 1 response they had learned and the corresponding List 2 response. The other *S*s, in the Unlearning (U) condition were warned that List 1 responses would interfere with List 2 learning, and advised to suppress or unlearn the List 1 response as quickly as possible. All *S*s were then trained on the second (A-C) list, and finally given an unpaced recall test of the MMFR variety to assess the amount of unlearning of List 1 responses.

Method

Materials.—The lists were those used in the A-B, A-C (or S_IR_N) paradigm of Dallett (1962), and are given in that paper. Twelve Glaze trigrams of high-association value were the stimuli, with adjectives as responses. Stimuli and responses were used in four different pairings (only three of which were used in Exp. I). The different pairings differed in pairing of stimulus and response:

[1] At the time Exp. I was performed, K. M. Dallett was a Predoctoral Research Fellow, on Fellowship MF12-379 from the National Institute of Mental Health, United States Public Health Service. Experiments II and III were supported by Research Grant MH 06589-01 from the same agency.

TABLE 1

E, PRESENTATION RATE, AND DEGREE OF LEARNING IN EXP. I–III

Exp.	*E*	Rate	Cond.	Degree of Learning	
				List 1	List 2
I	Dallett	2:3	M, U M, U	5 trials 15 trials	1 perfect trial 1 perfect trial
II	D'Andrea	2:2	M, U C	12 trials 12 trials	12 trials —
III	D'Andrea	3:3	M, U	9 trials	9 trials

Note.—Conditions: M = Mediation, U = Unlearning, C = Control (List 1 only).

in all of these pairings the pairing of responses between lists was the same. Thus, referring to the lists given by Dallett (1962), AGILE was always the List 2 response corresponding to (i.e., paired with the same stimulus as) UNKIND on List 1, BELOVED corresponded to PETTY, etc. For each *S*, lists were presented in four different orders to reduce serial learning.

Procedure.—Each *S* was seated before the memory drum and given conventional instructions for paired-associate learning. He was not required to respond on the first presentation, but was to attempt to anticipate each response on trials after the first. It was made clear that errors would not be penalized. The first list was then presented for a fixed number of trials, the exact number of trials and the rate of presentation differing from experiment to experiment (Table 1). Following List 1 learning, the drum was stopped and *S* was given one of two sets of instructions. For the M condition, the instructions were:

> Now I would like you to learn a second list. The syllables will be the same but the words are different. Try to use the words you have already learned to help you in learning this new set of words. If you can, try to find some way in which you can relate them, or some way in which the word on the first list helps you remember the word on the second list. For instance, if the first-list word was *speedy* and the new word is *lacy*, you might want to think of a girl in a lacy dress driving a sports car—or you might just want to remember that *lacy* goes with the same syllable that *speedy* went with before. Some people claim that the more ridiculous the association you form between the words, the easier they are to learn and remember. Exactly what you do is, of course, up to you. Try to learn this second list as quickly as possible. But try to make what you have already learned on the first list help you as much as possible.

The *Ss* in the Unlearning groups were told:

> Now I would like you to learn a second list. The syllables will be the same but the words are different. Try to forget the words you have already learned to these syllables: if you do not they will interfere with learning this second list. Most people who have trouble learning this second list say it is because words from the first list keep intruding; try to prevent this. Exactly what you do is, of course, up to you. Try to learn this second list as quickly as possible. But be on the lookout for interference from the first list and try to suppress or forget the first-list words as soon as you can.

Two aspects of these instructions require comment. First, while Barnes and Underwood (1959) emphasized mediation "chains" of the A-(B)-B′ variety, the present instructions also allow for "rings" in which the two responses may be separately related to the stimulus and to each other, and for a "branch" structure in which the B and C responses are learned as compatible responses to the same stimulus. For purposes of constructing the instructions, all of these were treated as mediators, or as alternatives to B-C incompatibility. Secondly, the instructions include an escape clause for *S*; this was inserted to prevent an overconscientious *S* for whom the instructions might be disadvantageous, from attempting to *prevent* learning which occurred spontaneously or by other means. In short, the instructions were designed to influence *S*'s behavior but not to prescribe it. What *S* does as a result of such an instruc-

TABLE 2

MEANS AND *SD*s OF PERFORMANCE IN LIST 2 LEARNING AND SUBSEQUENT LIST 1 RECALL

Exp.	Cond.	List 2 *M*	List 2 *SD*	List 1 Recall *M*	List 1 Recall *SD*
I	M-5	11.55	4.07	3.44	3.70
	U-5	12.67	5.53	3.39	2.64
	M-15	10.27	5.11	9.17	2.71
	U-15	11.16	4.63	7.11	3.98
II	M	79.35	16.28	7.25	3.54
	U	72.95	23.25	5.70	2.83
	Control	—	—	10.10	1.78
III	M	61.60	16.65	9.05	2.52
	U	64.40	17.73	7.60	3.37

Note.—The List 2 learning measures are trials to criterion (Exp. I) or total correct anticipations (Exp. II, III).

tional stimulus is a theoretical question requiring more than an assumptive answer.

The second list was then presented to a criterion of one perfect trial (Exp. I) or for a fixed number of trials (Exp. II, III) at the same rate used for List 1. These details appear in Table 1.

Immediately after List 2 learning, *S*s were given an MMFR test of recall, in which they were handed a dittoed list of the stimulus syllables, each with two vertically arrayed blank spaces on which *S* was asked to enter the two responses. No time limit was placed on this test of recall. The syllables were presented in three different orders to different *S*s in Exp. I, with five different orders being used in Exp. II and III.

Finally, each *S* was handed a sheet containing all 12 stimuli each paired with the appropriate set of two responses. On this sheet *S*s were asked to indicate any "associations, mnemonic devices or 'tricks'" which had been used in learning any of the items. Most *S*s seemed to have no difficulty in filling this out, and no attempt was made to force answers.

Subjects and interexperimental differences.—Most of the differences between experiments are outlined in Table 1. The *S*s in all three experiments were students in introductory psychology classes who participated as part of a class requirement, and were assigned to conditions in rotation prior to their appearance in the laboratory. Experiment I was carried out at the University of California, Berkeley, in the summer of 1961 using a Phipps and Bird memory drum. In Exp. I there were 18 *S*s in each of the four groups, most, but not all, being naive with respect to learning in the laboratory. In Exp. I it was necessary to replace two *S*s who were not trying to anticipate overtly on early trials, and a third *S* was discarded for *E* error.

Experiments II and III were carried out at the University of California, Los Angeles, in 1962 and 1963, using a Stowe (459B) memory drum. Experiment II contained, in addition to the M and U conditions, a control group which learned the first list, was given the Block Design and Digit-Symbol subtests of the WAIS for the time passed by other groups in List 2 learning, and was tested for List 1 recall. In Exp. II and III each group contained 20 *S*s, none of whom had prior experience in learning experiments. In Exp. II one *S* had to be discarded for *E* error.

RESULTS

List 1 learning.—Analyses of variance were used to compare the M and U conditions in each experiment separately. The groups were treated identically during first-list learning, and in none of the experiments did significant differences appear, the F ratios for Exp. I and III being less than 1.00, and for Exp. II, $F\ (2, 57) = 1.243, p > .05$. Intralist and Other errors were also examined. In List 1 learning as in all of the analyses to be reported, examination of overt errors failed to yield findings of significance, and will not be further discussed.

List 2 learning.—Analyses of variance comparing trials to criterion in the M and U conditions of Exp. I and total correct anticipations in Exp. II and III yielded, respectively, $F\ (1, 68) = .761$, $F\ (1, 38) = 1.420$, $F\ (1, 38) = .294$, all p's $> .05$. The means and *SD*s of these List 2 performance measures appear in Table 2, where it is apparent that even the direction of nonsignificant differences is not consistent. Differential instructions do not appear to have influenced overall performance level in List 2 learning. However, this does not mean that

learning proceeded identically in M and U conditions. Closer examination of the data indicates that *S*s given mediation instructions tended to do best on List 2 items corresponding to well-learned List 1 items, while *S*s given unlearning instructions show no such tendency. In order to make these comparisons, the 12 items on each *S*'s first list, and the 12 items on his second list, were separately rank ordered according to the number of correct anticipations each item had obtained in learning. The two sets of ranks for each *S* were used to obtain the sum of squared differences between the ranks for corresponding List 1 and List 2 items. One such score was obtained for each *S*, expressing the relation between the difficulty or degree of learning of corresponding items on the two lists. This score (ΣD^2) is an index of relationship which can be transformed into the Spearman rank-order coefficient of correlation. An analysis of variance of these scores was carried out, using M-U as one variable and Comparisons as the other variable. Experiment I furnished 5-trial and 15-trial comparisons while Exp. II and III furnished one M-U comparison each. The analysis revealed a significant M-U difference, $F\,(1, 145) = 9.341$, $p < .001$, no difference among comparison means, $F\,(3, 145) = 1.003$, and a significant interaction, $F\,(3, 145) = 4.397$, $p < .01$. Individual t tests were used to test M-U differences in each of the four comparisons. The results of these t tests, together with the correlation coefficients corresponding to the mean ΣD^2 for each group, appear in Table 3. It will be seen that except for the 5-trial groups of Exp. I, the M groups show consistently higher mean correlations that the U groups, the difference being statistically significant in two of the four comparisons. The conditions for which this effect is not significant are the 5-trial conditions of Exp. I, in which *S*s have not learned enough of List 1 to provide stable mediators or stable indexes of relative item difficulties, and the conditions of Exp. II in which the presentation rate was slightly faster than the rates used in Exp. I and III. It is worth noting that there is no evidence to suggest that usually, in the A-B, A-C paradigm, the best-learned List 2 items should correspond to the most poorly learned List 1 items, although a simple response-competition assumption might lead to such a prediction. On the contrary, the data of Underwood (1951) indicate a positive correlation, and there are reasons (e.g., common stimuli) for expecting this to be generally true. While the differences in these interlist item-difficulty correlations are not large, they are consistent with the assumption that *S*s were responding differentially to the different instructions which preceded List 2 learning; under such an assumption Mediation *S*s should focus first on items for which well-learned List 1 mediators are available, while Unlearning *S*s should ideally concentrate on items for which there is least unlearning to be done.

TABLE 3

Rank-Order Correlations of List 1 and List 2 Item Difficulty

Exp.	M	U	t
I-5	.144	.266	1.14
I-15	.299	−.068	4.19**
II	.238	.074	1.62
III	.324	.117	2.46*

Note.—The t ratios for Exp. I have $df = 34$, those for Exp. II and III, $df = 38$.
* $p < .05$.
** $p < .01$.

Recall.—Second-list recall was virtually perfect in every condition, and

TABLE 4

CLASSIFICATION SCHEME USED IN TABULATING MNEMONICS

Category	Definition	Example
S	Mentions only the stimulus.	DEP-HEAVY-COMPLETE: "DEP = *deep*."
S-1, S-2	Separate associations linking stimulus and List 1 response (S-1) or stimulus and List 2 response (S-2).	DEP-HEAVY-COMPLETE: "DEP = *deep*, which suggested HEAVY: On List 2, COMPLETEly DEPendent" "DEP-HEAVY both have E, while DEP-COMPLETE share P."
S-1-2	Single mnemonic incorporating the stimulus and both responses, *or* a single stimulus interpretation which is described as appropriate to both responses.	DEP-HEAVY-COMPLETE: "DEP = *deep*, and things which are HEAVY and COMPLETE are likely to sink, suggesting *deep*."
1-2	List 1 response linked to List 2 response without explicit mention of stimulus relation to List 2 response. Can be part of S-1, 1-2 chain.	DEP-HEAVY-COMPLETE: "Things which are COMPLETE might also be HEAVY."

will not be discussed. The Means and *SD*s for List 1 recall appear in Table 2. The Exp. II Control condition suggests that retroactive inhibition has occurred, and is of no further importance. The differences for the 5-trial groups in Exp I are obviously insignificant. Analyses of the M-U differences in the 15-trial groups of Exp. I, and Exp. II and III yield differences which are not statistically significant when each experiment is considered separately: for Exp. I, $F(1, 34) = 2.996$; for Exp. II, $F(1, 38) = 2.342$; for Exp. III, $F(1, 38) = 2.382$. If, however, all of the M-U differences (including the 5-trial groups of Exp. I) are included in a single analysis, it becomes clear that the M groups recalled slightly more (a mean of 7.28 items) than the U groups (5.99 items) with $F(1, 144) = 6.186$, $p < .05$. Tabulation of List 1 responses as a function of the number of times they were correctly anticipated in learning suggests that the superiority of the M groups occurs for items at an intermediate level of difficulty (or degree of learning), and not for the very strongest or the very weakest items. The necessity of a combined analysis to achieve significance at the .05 level indicates that the overall differences are quite small in relation to intragroup variability.

Reported associative aids.—Following the test of recall, *S*s reported associative aids used in learning. While such retrospective reports need have nothing essential to do with the learning process, they still represent a facet of verbal learning which needs to be better understood. The reports obtained in each experiment were classified, using the categories outlined in Table 4. These categories were designed after the data from Exp. I had been examined, and were intended to capture all reports of potential mediators which linked List 1 items with List 2 items. Operating on the assumption that the number of such mediators was the essential datum of interest, in the final tabulation dubious cases of mediation were included as instances of mediation. Treating dubious cases as instances of nonmediation reduces the number of reported mediators but changes no conclusions. Each of the authors scored Exp. II and III independently, finding the same results in every in-

TABLE 5

MEAN NUMBER OF MNEMONICS REPORTED AFTER LEARNING

Exp.	Cond.	Category				
		S	S-1	S-2	S-1-2	1-2
I	M-5 U-5	.22 .55	2.50 2.50	5.83 5.05	1.55 1.16	.55 .22
	M-15 U-15	.11 .33	3.22 5.16	3.00 5.55	2.50 2.17	3.38 .17
II	M U Control	.05 .25 .45	2.40 3.40 3.40	2.75 4.15 —	1.55 .65 —	2.00 .10 —
III	M U	.45 .70	2.60 3.20	1.75 4.25	3.35 1.00	3.00 .10

Note.—See Table 4 for explanation of scoring categories.

stance but differing in the number of mediators found; these differences could be resolved after discussion of the criteria each had adopted. (The tabulation presented in Table 5 is that of Dallett.) In examining Table 5, the categories of interest are S-1-2, and 1-2, which include reports of associations which link or incorporate both List 1 and List 2 responses. The S-1 and S-2 categories, on the other hand, represent independent mnemonics. As an example, consider Cond. M-5 of Exp. I. An average of .22 mnemonics per *S* involved only the stimulus, and 2.50 + 5.83 were single-list mnemonics linking stimulus and response, for a total of 8.55 out of 24 possible single-item mnemonics per *S*. The mediating responses included 1.55 in the S-1-2 category (involving both S-1 and S-2), and .55 in the 1-2 category which may have been relevant only to List 2, or which may have accompanied an S-1. Hence, each *S* gave, on the average, mnemonics for 12.20 out of his 24 items, a rate close to 50%. The one-list Control of Exp. II suggests that List 2 learning does not seriously distort the rate of reporting List 1 mnemonics, which varies from about 25% to about 45%. Overall rates undoubtedly depend upon the materials, instructions, and other conditions of testing, and only relative rates are of interest here. It will be seen that mediating responses are more frequently reported in the M groups in each of the eight possible comparisons, with a compensatory surplus of independent mnemonics in the U groups in six of the eight comparisons involving S-1 and S-2 and in each of the four comparisons involving *S* alone.

As the examples in Table 4 might suggest, many of these reports have a flavor of rationalization about them, and may well have arisen in an attempt to "explain" (after the fact) why a given item was an easy one—or they may represent *S*'s attempt to indicate that he is an "intelligent" rather than a "rote" learner. At the same time, the data are consistent with the assumption that *S*s in the two groups were responding to the differential instructions, and that *S*s in the M groups were especially likely to be aware of potential relationships linking their first and second lists.

Discussion

While *S*s' apparent compliance with instructions to mediate or to unlearn did not affect the degree of List 2 learning at all, it did have a small effect on List 1 recall. This effect might have arisen in two ways: one way is through the difference in relations between List 1 and List 2 item strengths, and the other way is through the assumed occurrence of different numbers of mediating responses in the two experimental conditions. In the item-strength explanation, it is assumed that List 1 items weaker than or equal to their corresponding List 2 items are especially susceptible to interference: in the M groups the strong List 2 items tended to correspond to strong List 1 items, while in the U groups there is no such relationship. Hence, in the M groups much of the interference potential of strong List 2 items was wasted against items which were too well learned to be easily forgotten, and the total amount of retroaction was thereby reduced. Unfortunately, the interlist correlations are quite small, rendering this a very subtle effect which would be hard to verify in detail. The item analyses suggest that it is items of intermediate strength which differ. This would seem to support an explanation based upon the assumption that *S*s in the mediation groups attempt to use List 1 responses as mediators. This gives extra practice in recalling certain List 1 responses, and this extra practice is reflected in greater resistance to interference. The items affected are those weak enough to profit from the extra practice, and yet strong enough to be recalled and used as mediators in List 2 learning.

The results are also relevant to two recent experiments which have shown that retroaction can be affected by instructions given prior to List 2 learning (Postman & Stark, 1962; Schwartz, 1963). In each of these experiments, the instructions related to the forthcoming test of List 1 recall, Postman and Stark (1962) finding that *S*s warned of List 1 recall did better than those not warned, and Schwartz (1963) finding that *S*s warned that List 2 would make List 1 recall harder did better than *S*s warned that List 2 would make List 1 recall easier. The present experiments differ in that no reference was made to List 1 recall, and also differed in that the test of List 1 recall was unpaced. Otherwise, the present results agree with those of Postman and Stark (1962) and Schwartz (1963); a difference in recall appeared with no difference in overall List 2 performance. The mediation hypothesis offers one mechanism through which such effects can be explained.

References

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Dallett, K. M. The role of response similarity in proactive inhibition. *J. exp. Psychol.*, 1962, **64**, 364–372.

Dallett, K. M. Proactive and retroactive inhibition in the A-B, A-B′ paradigm. *J. exp. Psychol.*, 1964, **68**, 190–200.

Postman, L., & Stark, K. Retroactive inhibition as a function of set during the interpolated task. *J. verbal Learn. verbal Behav.*, 1962, **1**, 304–311.

Schwartz, H. A. Influence of instructional set and response frequency on retroactive interference. *J. exp. Psychol.*, 1963, **66**, 127–132.

Underwood, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.

(Received February 20, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 467–475

EFFECTS OF DELAYED AUDITORY FEEDBACK ON MORSE TRANSMISSION BY SKILLED OPERATORS

AUBREY J. YATES [1]

University of Western Australia

3 experiments were carried out on the effects of delayed auditory feedback (DAF) on the transmission of letters by 6 skilled Morse code operators. In the 1st experiment, random sequences of 8 letters were transmitted under 3 conditions (each letter transmitted separately, continuously at preferred rate, or continuously as fast as possible) at a constant delay time of 180 msec. In the 2nd experiment, the same sequences were transmitted as fast as possible at delays varying from 30 to 300 msec. as well as under no delay. In the 3rd experiment, meaningful material was transmitted as fast as possible under a delay of 180 msec. The results showed that DAF produces a great increase in the number of errors made; that the errors almost always involve an additional symbol or symbols; and that letters involving 3 or 4 symbols produce many more errors than letters involving 1 or 2 symbols.

A considerable body of research has now accumulated concerning the effects of delayed auditory feedback (DAF) on speech (Yates, 1963a, 1963b). It is also known, however, that DAF may affect other kinds of behavior in similar ways. The disturbing effect of DAF on rhythmic tapping was noted by Lee (1951) in his original study and was subsequently confirmed by Chase, Harvey, Standfast, Rapin, and Sutton (1959; Chase, Rapin, Gilden, Sutton, & Guilfoyle, 1961). In both these studies it was noted that, under DAF, the key was tapped harder, held down longer, more taps were given than asked for, and the pauses between taps were lengthened. More recent studies have extended these findings. Chase, Sutton, Fowler, Fay, and Ruhm (1961) found that disturbance in tapping rhythm was produced by a feedback level as low as 10 db. above sensation level. Chase, Rapin, Gilden, Sutton, and Guilfoyle (1961) showed that rate and intensity of tapping were affected by decreased auditory feedback (where *S* had been trained to tap at a specific rate, and with a specific amount of pressure). DAF was found to produce much more disturbance in a complex tapping task than in the specific rate/pressure task.

It is well known that the optimal delay (at high feedback intensity levels) for producing maximal speech disturbance is about 180 msec. However, very little is known at present concerning the explanation of this phenomenon, except that the time relationships between the output and its associated feedback are probably very important. Unfortunately, the fundamental speech units involved in monitoring are not known and hence the analysis of speech disturbance is not very suitable for the elucidation of the basic time relationships between sensory reception, central processes, and motor behavior under conditions of DAF. The systematic investigation of complex tapping performance under normal and delay conditions may be more suited for this purpose, since the units involved (dots and dashes) are discrete and readily quantifiable in terms of errors and trans-

[1] Now at the University of New England, Armidale, N. S. W.

TABLE 1

LETTERS AND SYMBOLS USED IN EXP. I AND II

One Symbol	Two Symbols	Three Symbols	Four Symbols
E .	A .–	W .––	J .–––
T –	N –.	D –..	B –...

mission times. The experiments reported here have attempted to clarify these relationships by making use in the first two experiments of a restricted sample of the Morse code alphabet. In the first experiment, skilled operators transmitted letters involving one, two, three, or four symbols under a delay of 180 msec. at three rates of sending. In the second experiment, the effects of varying delay intervals were investigated. In the third experiment, the effect of DAF on continuous transmission of meaningful material was studied. The basic hypothesis was that the effect of delay would be greater, the longer the unit being transmitted, on the assumption that with very short units, the letter will have been transmitted before the delayed feedback is received, whereas with longer units, the message will be still in process of transmission when the delayed feedback commences.

PILOT STUDY

In a series of preliminary experiments, utilizing the materials described below, it was verified that *S*s did not make errors when transmitting under normal (no-delay) conditions while wearing headphones; nor did they make errors when masking white noise was presented through the headphones while they were transmitting, so that they could not hear the sound of the key. It was concluded that the absence of auditory feedback did not affect the accuracy of transmission.

EXPERIMENT I

Method

Subjects.—The *S*s were six employees of The Postmaster General's department. All were highly skilled operators, with many years of experience, and were accustomed to receiving and transmitting in noisy conditions over long distances.

Apparatus.—The *S* transmitted with a standard Morse key, the output of which, under the delay condition, was fed into a delayed feedback machine and returned to *S*'s ears through headphones after a delay of 180 msec. The amplified playback intensity at the headphones was set at 100 db., as indicated by a Dawe sound-level meter, Type No. 1400E, which was acoustically coupled to one ear phone of the headphone set. The coupling was by a sound column of ½ in. between the microphone and the headphone.

Materials.—The eight letters shown in Table 1 were used. It will be noted that there are four groups of letters, with two representatives each of letters involving one, two, three, and four symbols. Twelve lists were constructed, each containing 32 letters, i.e., each letter appeared four times. The order of the letters in each list was determined from tables of random numbers. Each list was printed on a separate sheet, with eight rows to each sheet. Thus, a different list was used with each of the 12 conditions described below.

Experimental design.—Three transmitting conditions were used. In the first condition, *S* transmitted each letter *separately*, i.e., with a slight pause between each letter. In the second condition, *S* transmitted *continuously* at his own *preferred* rate. In the third condition, *S* transmitted *continuously* as *fast as possible*. Each of the 3 conditions was replicated four times, the sequence of the 12 conditions being randomly determined for each *S*. Each *S* was tested individually.

Scoring.—An electrocardiograph pen recorder was used to produce a permanent record of *S*'s performance. No difficulties were encountered in scoring the records for errors and transmission time for individual letters was readily calculated.

Results

Table 2 shows the distribution of errors, summed over *S*s, for the three transmission rates, by letter. Several striking facts are apparent. Errors involving letters of three or four

symbols are nearly 15 times as common as errors involving letters of one or two symbols. Errors for four-symbol letters are twice as frequent as errors for three-symbol letters. For the three- and four-symbol letters, the errors for D and B (dash followed by dots) are twice as frequent as the errors for W and J (dot followed by dashes), respectively. Finally, an equal number of errors is produced in all three transmitting conditions, the pattern of errors for the eight letters being very similar for each condition.

Discussion

As noted in earlier experiments on less complex tapping tasks, it is clear that DAF has a strikingly disruptive effect on the accuracy of transmission of letters of the Morse code by highly skilled operators, who are accustomed to transmitting under difficult conditions. Further, as hypothesized, the effect of DAF is negligible provided the letters being transmitted involve only one or two symbols. A very significant increase in errors occurs, however, when the letters being transmitted involve three or four

TABLE 2

Error Scores for Each Letter under Three Transmitting Conditions: Exp. I

Letter	Transmitting Cond.			
	Separately	Preferred Rate	Fast as Possible	Percentage Error
E	1	0	1	1
T	0	1	0	0
A	4	1	0	2
N	5	4	2	4
W	11	11	10	11
D	23	25	19	23
J	17	22	25	22
B	37	39	40	40
Total	98	103	97	13

Note.—For each cell in the table, the maximum possible error score (summed over *S*s) was 96.

TABLE 3

Transmission Time (in Msec.) for Eight Letters under Three Conditions (180-Msec. Delay): Exp. I

Letters	Transmission Cond.		
	Separately	Own Rate	Fast as Possible
E	050	046	044
T	268	239	197
A	374	336	286
N	317	291	267
W	556	520	477
D	483	446	420
J	824	711	668
B	631	575	559

symbols. Furthermore, the severity of the effect of delay on the letters involving longer symbols appears to depend on whether the symbol consists of a dot followed by dashes or the reverse, with the latter being more susceptible to error. Examination of the individual records showed that this effect was clearly present in five *S*s for the letters J and B, and in three *S*s for W and D (the sixth *S* was unaffected by DAF).

These results, therefore, may appear at first sight to provide some support for the hypothesis that the transmission time of one- and two-symbol letters is so short that transmission of the signal has been completed before the feedback arrives back at the cortex and hence cannot be disrupted by it. (It would appear to be equivalent, therefore, to transmitting in the *absence* of feedback—as we have seen, skilled *S*s are capable of doing this without error.)

A more direct check of the hypothesis may be obtained, however, by calculating the transmission times for the eight letters used in this experiment. If the hypothesis is correct, then the transmission time for letters involving one and two symbols should be less than the delay time used in this experiment, i.e., 180 msec., while the transmission time for three- and four-symbol letters should be greater. The objective pen tracings of *S*'s performance enabled a check on transmission times to be made, since the

paper traveled at a constant speed. The transmission time was taken to start at the point where the pen began its upstroke on the first symbol and to terminate at the point where the pen began its downstroke on the last symbol of a letter. The transmission time for all letters *correctly transmitted* was calculated in this way with the results shown in Table 3 for all *S*s combined, a total of 2,007 responses being scored.

It will be noted that the instructions to transmit separately, at preferred rate, or as fast as possible, were successfully communicated to *S*, since for each letter there is a consistent increase in transmission speed over the three conditions. However, it is clear that the results offer no support whatever for the hypothesis, since all letters except E took longer than the delay time (180 msec.) to transmit. Since few errors were made on the letters T, A, and N, it follows that overlap between transmission time and delay time is not a sufficient condition for the disruption of performance in this particular task.

Experiment II

It is well known that there is a differential effect of DAF on speech at high intensities of feedback (Butler & Galloway, 1957). Black (1951) showed that there is a discrete stepwise increment in the effect at a delay of 60 msec., that between 60 and 180 msec. there is a linear relationship between increased delay and increased disturbance of speech, and that between 180 and 300 msec. delay the effect remains fairly constant. With delays longer than 300 msec., the effect gradually lessens and speech returns to approximately normal when the delay becomes greater than 500 msec. Little is known, however, of the effects of differential delay on tapping, although a recent study by Rapin, Costa, Mandel, and Fromowitz (1963) with children which used delays from 0 to 1,000 msec. in 100-msec. steps suggests that delays up to 1,000 msec. produce roughly the same amount of disturbance. Nothing is known of the differential effects of delay on *skilled* tapping performance. Experiment II was designed to throw light on this problem.

Method

The *S*s, apparatus, and materials were the same as those used in the first experiment, this experiment being carried out immediately after Exp. I.

Experimental design.—The 12 sets of 32 letters were randomly paired with 12 delay times, the randomization being carried out afresh for each *S*. Ten different delays were used (30–300 msec. by 30-msec. increments). One of the sets was transmitted under no delay prior to the delay conditions, while another set was also transmitted under no delay after the delay conditions. All sets of letters were transmitted under instructions to work as fast as possible and to transmit continuously.

Results

The number of errors for each letter under each delay is shown in Table 4. There appears to be a discrete increment in the number of errors as the delay time increases from 90 to 120 msec. Delay times greater than 90 msec. produce roughly the same number of errors up to the limit of 300 msec. used in this experiment. These results agree with those of Rapin et al. (1963) as far as they go. Letters involving three and four symbols produce many more errors than letters involving one or two symbols, being approximately six times as frequent. However, there is no suggestion in this experiment that Letters W and J produce fewer errors than D and B; if anything, there is a reverse tendency. In Exp. I, the delay used was 180 msec. Reference to the delay time of 180 msec. in Table 4 confirms the absence of this differential effect.

TABLE 4

ERROR SCORES FOR EACH LETTER AS A FUNCTION OF DELAY: EXP. II

Letters	Delay Time (in Msec.)													Percentage Error
	00	30	60	90	120	150	180	210	240	270	300	00	Total	
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T	0	1	0	0	0	0	0	0	1	0	0	0	2	1
A	0	0	2	0	1	0	0	1	0	1	1	1	7	2
N	0	0	0	0	4	0	1	1	5	4	3	0	18	6
W	0	1	3	2	2	5	6	4	5	6	5	0	39	14
D	0	0	0	0	3	3	5	1	7	5	3	0	27	9
J	3	0	4	0	5	5	8	6	5	9	6	0	51	18
B	0	1	0	1	5	3	9	7	8	6	5	0	45	16
Total	3	3	9	3	20	16	29	20	31	31	23	1	189	7

Note.—For each cell in the table, the maximum possible error score (summed over *Ss*) was 24.

Discussion

While these results in general confirm the findings of Rapin et al. (1963) regarding the effects of different delays on tapping, there are two features of the results which require comment. First, the absence of a differential effect for three- and four-symbol letters, according to whether they commence with a dot and end with dashes or the reverse, is not readily explained. There is no doubt that, allowing for the absence of errors at very short delay times, the number of errors produced is proportionately as great as in Exp. I. (Thus, at a delay of 180 msec., 29 errors resulted from one quarter the number of symbols used in Exp. I. This gives an adjusted error total of 116 which compares quite well with the 97 errors produced under the equivalent condition in Exp. I.)

Second, the relative lack of errors at the short delays confirms the finding in Exp. I that overlap of the transmitted and delayed signals is not a sufficient condition for the disruption of performance. If the one- and two-syllable letters *had* been transmitted faster than the delay time in Exp. I then the shorter delays used in Exp. II should have disrupted correct transmission of these letters. However, aside altogether from the results obtained in Exp. I, it is clear from Table 4 that the shorter delay times do not have any differential effect on transmission of the one- and two-syllable letters. This finding is also confirmed by an analysis of transmission time scores for correctly transmitted letters, as shown in Table 5. There are two interesting features of this table, however. First, it would appear that the frequent occurrence of incorrectly transmitted letters in the delay condition has no effect on the speed of sending of adjacent correctly transmitted letters since the individual letter and overall average times for the two no-delay conditions are not markedly different from those for the delay times. Second, the transmission times for each individual letter remain remarkably constant over the different delays, that is, correctly transmitted letters are sent at roughly the same speed whether the delay time is long or short.

EXPERIMENT III

This experiment was concerned with the effects of DAF on the accuracy of transmission of meaningful prose material. No previous studies have been published which investigate this problem.

TABLE 5

TRANSMISSION TIMES (IN MSEC.) FOR EIGHT LETTERS UNDER DIFFERENT DELAY TIMES: EXP. II

Letters	Delay Time (in Msec.)												*M*
	00	30	60	90	120	150	180	210	240	270	300	00	
E	045	044	046	043	042	046	045	040	048	042	041	042	044
T	161	166	161	172	179	178	171	165	179	168	154	152	167
A	260	265	270	283	267	283	259	252	269	254	255	238	263
N	244	247	243	257	253	260	251	248	255	258	240	245	250
W	446	442	440	431	437	452	448	423	438	417	422	430	436
D	379	386	393	396	400	403	388	399	411	393	378	383	392
J	628	628	603	610	596	622	640	592	613	621	580	574	609
B	499	528	530	549	550	566	528	530	544	546	519	514	534
M	333	338	336	343	341	351	341	331	345	337	324	322	337

Method

The *S*s and apparatus were the same as those in the previous experiments.

Materials.—Two sets of six sentences each were prepared. One set consisted of sentences whose word structure was relatively uncomplicated (e.g., "Sometimes I climb mountains in my dreams."), the other consisted of sentences whose word structure was more complex (e.g. "crystallography and physical chemistry") though it is clear that the distinction might hardly be significant to such highly skilled *S*s.

Experimental design.—Each *S* transmitted both easy and hard sentences (at the speed which he would employ under normal transmission conditions) under no-delay and delay conditions. The order of passages and of delay/no delay was randomized. The delay used was 180 msec. A total of 190 letters was contained in the six "easy" sentences; while there were 209 letters in the six "difficult" passages.

Results

Table 6 shows the frequency distribution of occurrence (summed over the six *S*s) of each letter contained in the "easy" and "hard" sentences, arranged in groups, according to whether the letters are represented by one, two, three, or four symbols. The table also shows the absolute number of errors (summed over *S*s) for each letter under delay and no-delay conditions, together with the percentage of errors for each letter for the delay conditions only (errors in the no-delay conditions being negligible).

Inspection of the table shows that relatively few errors are made on letters involving one or two symbols (with the exception of Letter M in the "difficult" condition) under delay, with a very marked increase in the number of errors for letters involving three or four symbols. There appears to be little difference between the "easy" and "hard" conditions in this respect. Under the no-delay conditions, the number of errors was negligible. Once again there appears to be little difference in the errors produced by three- and four-symbol letters. It is interesting to note that the letters K (–.–) and R (.–.) which are symmetrical, alone produced virtually no errors among the three- and four-symbol letters.

Discussion

The difference between the total number of errors for the delay and no-delay conditions in both the "easy" (187 as against 9) and the "difficult" (212 as against 8) sentences illustrates very strikingly the profound disruption caused by the delayed feedback. The results closely parallel the findings of Exp. I and II using nonmeaningful sequences of letters.

TABLE 6

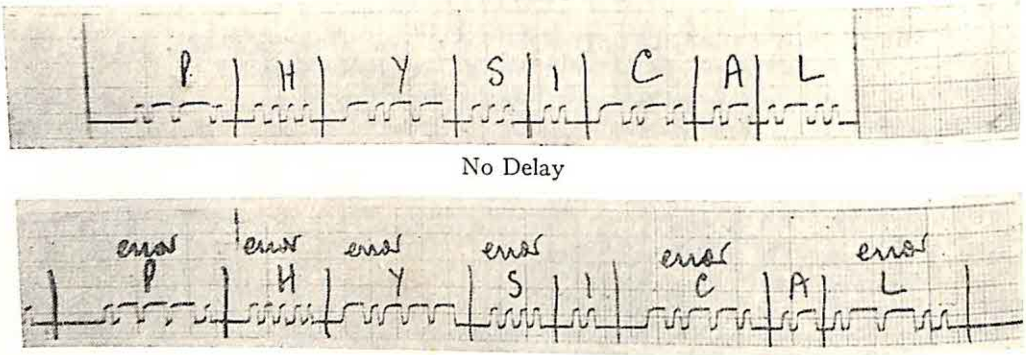
Error Scores for Each Letter in "Easy" and "Difficult" Material under No Delay and Delay Conditions: Exp. III

Letter	"Easy" Material					"Difficult" Material				
	No Delay		Delay			No Delay		Delay		
	Freq.	Errors	Freq.	Errors	%	Freq.	Errors	Freq.	Errors	%
E	138	0	128	11	8	144	0	144	5	3
T	90	0	80	0	0	138	0	138	2	1
A	60	0	56	0	0	126	0	126	2	2
I	72	0	70	7	10	120	0	120	7	6
M	54	0	54	3	6	30	1	30	7	23
N	96	0	94	1	1	84	0	84	5	6
D	24	1	22	8	33	30	0	30	9	30
G	42	1	40	16	38	6	0	6	2	33
K	6	0	6	0	0	6	0	6	0	0
O	84	1	78	26	31	78	0	78	35	45
R	78	0	72	3	4	78	1	78	2	3
S	90	2	85	24	27	90	1	90	29	32
U	48	0	44	8	17	18	0	18	2	11
W	12	1	11	3	25	6	0	6	2	33
B	48	0	43	15	31	12	0	12	6	50
C	18	1	16	3	17	66	2	66	16	24
F	12	0	11	2	17	18	0	18	3	17
H	48	1	43	18	38	66	1	66	36	55
L	42	0	40	15	36	42	0	42	11	26
P	48	0	45	14	29	54	0	54	14	26
Q	6	1	6	1	17	6	0	6	2	33
V	0	0	0	0	0	6	0	6	2	33
X	0	0	0	0	0	6	1	6	3	50
Y	24	0	23	9	38	24	1	24	12	50

Other Results

One of the most important findings only briefly mentioned so far is the fact that, without exception, the errors made did not affect the accurate transmission of adjacent letters, whether they were transmitted correctly or not. In other words, there was no carry-over of an error on one letter to the next. All *S*s continued to transmit at the same rate under delay as they had under no delay (Table 5). They appeared to be quite unaware of the very large proportion of errors which were being made under delay on the three- and four-symbol letters. When questioned after completion of testing, they admitted having difficulty with some letters but did not realize the extent of the errors. The types of errors made are clearly illustrated in Fig. 1 which shows the transmission of the word "physical" under no delay and delay. Transmission of the letters "i" and "a" is unaffected by the errors made on the other letters. The time data in Table 5 confirm this conclusion since these correctly transmitted letters were, of course, embedded in sequences of incorrectly transmitted letters.

Qualitative analysis of the types of errors proved difficult. Inspection of the data indicated clearly that an error almost without exception involved the transmission of an addi-



No Delay

Delay

FIG. 1. Types of errors occurring under delayed auditory feedback.

tional symbol, rather than omission of one of the symbols, of a letter. However, although casual inspection seemed to reveal that, for example, an extra dot was added in letters beginning with a dash and ending with dots, closer examination revealed many instances where it seemed possible that instead of adding an extra dot at the end of a letter, *S* might be inserting a shortened dash immediately following the first dash. In terms of the time relationship between transmitted response and perception of the feedback, a demonstration of which of these two possibilities was in fact occurring, becomes crucial. However, the interpretation of the transmission changes under delay would also depend in part on the characteristic mode of transmission of correctly sent letters under no delay. That is, in the case of the letter B (–...), for example, it is possible that some *S*s transmit a dash followed by three dots of equal transmission time, whereas other *S*s transmit a dash followed by three dots of unequal (or progressively diminishing) transmission time. Detailed analysis of the data from the first and third experiments supported this interpretation but the results are not reported in detail in this paper.

SUMMARY AND CONCLUSIONS

Three experiments were carried out to investigate the performance of skilled Morse code operators when the auditory feedback resulting from the transmission of symbols was delayed. In the first experiment random sequences of eight letters of the Morse code alphabet were transmitted under a constant delay of 180 msec. while *S* was transmitting either each letter as a separate unit, or continuously at own rate, or continuously as fast as possible. The principal results indicated:

1. Many more errors were made on letters involving three or four symbols as compared with those involving one or two symbols.
2. For the three- and four-symbol letters, those beginning with a dash and ending with a dot produced twice as many errors as those beginning with a dot and ending with a dash.
3. An equal number of errors was produced in all three transmitting conditions (separate, preferred rate, and fast as possible), with the results indicated above being found in all three conditions.

In the second experiment, the same random sequences of letters were used and *S* transmitted these sequences under 10 different delay times (30–300 msec. by 30-msec. steps) in addition to two control sequences under no delay. In all conditions, instructions were to transmit as fast as possible. The principal results indicated:

1. A discrete increase in the number of errors at about 120 msec. delay.
2. An equal effect on errors of delay times greater than 120 msec. up to the limit of 300 msec. used.

3. Confirmation of the findings in the first experiment that three- and four-symbol letters are much more subject to error under delay than one- and two-symbol letters.

4. Failure to confirm the finding in the first experiment that letters beginning with a dash and ending with dots produce more errors than letters beginning with a dot and ending with dashes.

In the third experiment the same *S*s transmitted meaningful material (categorized arbitrarily as "easy" or "difficult") under no delay and a delay of 180 msec. The principal results confirmed those found in the first experiment with the exception of the finding concerning the order of dots and dashes.

REFERENCES

BLACK, J. W. The effect of delayed sidetone upon vocal rate and intensity. *J. speech hear. Disord.*, 1951, **16**, 56–60.

BUTLER, R. A., & GALLOWAY, F. T. Factoral analysis of the delayed speech feedback phenomenon. *J. Acoust. Soc. Amer.*, 1957, **29**, 632–635.

CHASE, R. A., HARVEY, S., STANDFAST, S., RAPIN, I., & SUTTON, S. Comparison of the effects of delayed auditory feedback on speech and keytapping. *Science*, 1959, **129**, 903–904.

CHASE, R. A., RAPIN, I., GILDEN, L., SUTTON, S., & GUILFOYLE, G. Studies on sensory feedback: II. Sensory feedback influences on keytapping motor tasks. *Quart. J. exp. Psychol.*, 1961, **13**, 153–167.

CHASE, R. A., SUTTON, S., FOWLER, E. P., FAY, T. H., & RUHM, H. B. Low sensation level delayed clicks and keytapping. *J. speech hear. Res.*, 1961, **4**, 73–78.

LEE, B. S. Artificial stutter. *J. speech hear. Disord.*, 1951, **16**, 53–55.

RAPIN, I., COSTA, L. D., MANDEL, I. J., & FROMOWITZ, A. J. Effect of varying delays in auditory feedback of keytapping of children. *Percept. mot. Skills*, 1963, **16**, 489–500.

YATES, A. J. Delayed auditory feedback. *Psychol. Bull.*, 1963, **60**, 213–232. (a)

YATES, A. J. Recent empirical and theoretical approaches to the experimental manipulation of speech in normal subjects in stammerers. *Behav. Res. Theor.*, 1963, **1**, 95–119. (b)

(Received February 20, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 476–482

ABOLITION OF THE PRE BY INSTRUCTIONS IN GSR CONDITIONING [1]

WAGNER H. BRIDGER AND IRWIN J. MANDEL

Department of Psychiatry, Albert Einstein College of Medicine

80 human Ss divided into 2 groups of CRF and 2 groups of 25% PRF were used in a GSR classical conditioning paradigm to test the discrimination hypothesis as an explanation of the PRE. Removal of the shock electrodes and instructions as to the onset of extinction was assumed to provide equalization of discriminability between the acquisition and extinction series for 1 PRF and 1 CRF group. A postexperimental inquiry for these Informed groups provided a test of discriminability independent of resistance to extinction. The results demonstrated the presence of the PRE in the Noninformed groups and its abolition in the Informed. The presence of a conditioned GSR during extinction in the Informed groups combined with the abolition of the PRE suggests a 2-component CR characterized as a simple conditioned CR and a mediated CR.

The phenomenon of partial (intermittant) reinforcement leading to greater resistance to extinction than continuous reinforcement has become a firmly established empirical law in psychology (at least above the phyletic level of the fish). A frequently invoked explanation of this partial-reinforcement effect (PRE) has been that provided by the discrimination hypothesis (Mowrer & Jones, 1945). This hypothesis maintains that the rate of extinction of a response is inversely related to the similarity between acquisition and extinction conditions. There have been many attempts to refute the hypothesis, most notable of which have been Weinstock (1954), Wilson, Weiss, and Amsel (1955), and most recently Theios (1962) and Jenkins (1962).

The attacks on the discrimination hypothesis have, for the most part, been so structured as to depend upon *E*'s evaluation of the similarity between acquisition and extinction rather than the similarity as perceived by *S*. The *E*'s evaluation of discriminability has frequently been directed to the relative abruptness of the transition between acquisition and extinction. The discrimination hypothesis, however, need not be construed so as to be solely dependent on the abruptness of the local transition. Bitterman, Feddersen, and Tyler (1953) have maintained that discriminability may be a function of the degree of congruence between the total patterning of the acquisition and extinction series. Discrimination, however, does not lie either in the patterning of the stimuli or in *E*, but in *S*. An adequate evaluation of the explanatory power of the discrimination hypothesis has not been presented due to a failure to obtain a measure of *S*'s ability to discriminate between acquisition and extinction which is independent of the very response process to which the hypothesis is directed, i.e., resistance to extinction.

[1] This study was supported in part by Research Grants K3-MH-4177 and MH-06282 from the National Institute of Mental Health. The authors wish to thank James Block and William Swayze for their technical assistance in the use of a computer in both data reduction and analysis. A brief report of this experiment was presented at the Eastern Psychological Association, April 1964.

The need for such an independent evaluation has been noted by Bitterman et al. (1953). According to the discrimination hypothesis, therefore, if two groups of *S*s, one partially reinforced (PRF) and one continuously reinforced (CRF), could equally discriminate between an acquisition and an extinction series, the PRE should be abolished. In the present experiment it is assumed that informing *S*s as to the onset of extinction equalizes the discriminability for the two groups. An independent measure of the effect of the extinction information on *S*'s ability to discriminate between acquisition and extinction was obtained by means of a postexperimental inquiry.

Method

Subjects.—The *S*s were 116 male paid volunteer college students 18–25 yr. from various schools in the New York City area. Three *S*s were eliminated due to experimental error or equipment malfunction, 7 for failure to show a GSR to the UCS, 2 for having a negative GSR (an increase in skin resistance) to the UCS, 4 refused to continue because the UCS was too strong, and 20 were disqualified on the basis of a postexperimental inquiry to be described later. A total of 80 *S*s remained.

Apparatus.—The GSR was measured via two Ag/AgCl nonpolarizing sponge electrodes described by O'Connell and Tursky (1960) and evaluated by O'Connell, Tursky, and Orne (1960). A constant 100 μamp. dc current was applied to the electrodes via a Wheatstone bridge. Skin resistance was recorded on one channel of a transistorized eight-channel Type R Offner dynograph running at a paper speed of 5 mm/sec. The gain settings varied from .5 mv/cm to 10 mv/cm depending on the magnitude of *S*'s response. The UCS was a 500-msec., 23-v., 1.53-ma. (re: 15,000 ohms) ac shock delivered to *S*'s right shin bone via two $1\frac{1}{2} \times 3$ in. electrodes placed approximately 4 in. apart over electrode jelly. In a previous experiment (Bridger & Mandel, 1964) this shock level was uniformly reported by all *S*s as very painful. A differential conditioning procedure consisting of one positive (CS+) and four negative (CS−) stimuli was used. Each CS was a 500-msec. flash of one of five 25-w. bulbs mounted horizontally on a Masonite panel with centers 6 in. apart. With *S* lying supine, the panel was placed 7 ft. up on a wall approximately 10 ft. from *S*'s head. Onset and duration of both CS and UCS were controlled by Hunter timers, with the onset of the UCS occurring simultaneously with the offset of the CS. Onset and duration of both stimuli were recorded on separate channels of the dynograph.

Design.—Four groups of *S*s were used, two PRF and two CRF. At the onset of extinction, one PRF and one CRF group were informed as to the onset of extinction (PRFI, CRFI) as described in the procedure section. The two remaining groups were left uninformed (PRFN, CRFN). A single trial consisted of the presentation of each of the five lights in a random order. The extreme left-hand light was always designated as the light to be followed by shock (CS+) with none of the remaining lights ever followed by shock (CS−). The experiment had three parts: (*a*) a 5-trial adaptation series during which *S* was never shocked; (*b*) a 20-trial acquisition series during which the two CRF groups were shocked at each presentation of the CS+ (100% reinforcement) while the two PRF groups were shocked on Trials 1, 4, 11, 12, and 20 (25% reinforcement); (*c*) a 30-trial extinction series. The interval between successive CS presentations both within and between trials varied from 10 to 20 sec.; in general, a CS light was presented when *S*'s skin resistance either remained unchanged or was increasing for a few seconds.

Procedure.—The *S*s were randomly assigned to one of the four groups. The *S* was led into a well-lit sound-deadened experimental room isolated from an adjoining instrument room. The *S* was informed that the purpose of the experiment was to measure some of his physiological responses. The right palm and ventral surface of *S*'s right wrist were cleaned with an acetone solution. A 2-in. square of adhesive tape with a 1-in. diameter hole was placed on his palm. The palmar surface showing through the hole in the tape and the ventral surface of the wrist were coated with Cambridge electrode jelly. The GSR electrodes were placed over the jelly and held by rubber straps. The *S* was told their function. The shock electrodes were attached and their purpose explained as *S* lay supine on a mattress placed on a table. He was then informed that the experiment would contain three parts: (*a*) a period when his responses to the lights alone, as they flashed on and off, would be measured; (*b*) a period when the lights would be paired with

electric shock; (*c*) a final period during which his responses to the lights alone would again be recorded. The *S* was assured that nothing would occur during the experiment about which he had not previously been informed. The adaptation series was then presented. At the end of the series, *E* returned to *S*'s room and reiterated that there would be two more parts to the experiment; one part during which shocks would be given and a second part during which no shocks would be given. The CS+ light was then designated and *S* told that shocks "would be associated with" the light. The *S*s in the two Noninformed groups were told that *E* would not return until the experiment had been completed. The acquisition series was then presented and, for the Noninformed *S*s, the extinction series followed immediately without interruption. For the Informed *S*s, *E* returned at the end of the acquisition series, removed the shock electrodes, wiped *S*'s leg clean of electrode jelly, and reassured *S* that no further shocks would or could be given. The *S* was told that the purpose of the final part of the experiment was to see what effect the shock had on his responses to the lights presented alone. The extinction series was then presented.

Postexperimental inquiry.—To provide an independent measure of *S*'s ability to discriminate between the acquisition and extinction series, a postexperimental inquiry was conducted. The Informed *S*s were given a questionnaire containing the following question, "After you were told that no more shocks would be given and the electrodes were removed, how sure were you that you would not be shocked?" Below this statement appeared a horizontal 10-cm. line at one end of which was the phrase, "Positive no shocks"; at the other end, "Positive more shocks." The *S* was instructed to indicate by a vertical line his degree of belief or disbelief. Any *S* who did not mark the absolute end of the line indicating "Positive no more shocks" was eliminated from the sample. Twenty records were discarded by this procedure. A further interview revealed that most of these *S*s felt they might have continued to be shocked through the GSR electrodes.

Response measure.—By means of a calibration template for each sensitivity setting, it was possible to measure skin resistance at any point to the nearest 10 ohms. The final score assigned to a response was the change in conductance (ΔC) expressed in micromhos according to the formula $(1/Ra - 1/Rb)\ 10^6$, where Rb is the resistance at the moment of response and Ra is the resistance at the peak of a response. In order for a resistance change to be counted as a response, the decrease in skin resistance had to begin not less than 1.5 nor more than 4 sec. after the onset of any CS.

RESULTS

Two methods were used to analyze resistance to extinction, (*a*) Anderson's (1963) shape-function method, and (*b*) an extinction criterion measure.

Shape-function method.—Anderson's formula for shape-function scores which is designed to equate for differences in acquisition level, follows:

$$f(n) = R(\infty) - R(n)/R(\infty) - R(1),$$

where for each individual:

$f(n)$ = shape-function score on Extinction Trial n.

$R(\infty)$ = asymptotic extinction response level.

$R(n)$ = observed response on Extinction Trial n.

$R(1)$ = estimated response on Extinction Trial 1.

$R(\infty)$ was presumed to equal zero; i.e., given a sufficient number of trials, it was assumed that all *S*s would cease responding. The formula then reduces to a simple ratio score: $f(n) = R(n)/R(1)$. The $R(1)$ term represents an estimate of each *S*'s acquisition level before the onset of the extinction process. The most unbiased and reliable estimate of $R(1)$ was taken to be the mean acquisition level reached by each *S*. However, due to the short CS-UCS interval (500 msec.) the mean response level to the CS+ for the CRF groups could not be obtained. Examination of the acquisition performance of the two PRF groups revealed a high correlation (Pearson $r = .81$, $N = 40$) between each *S*'s mean response to the CS+ (MCR+) on the 15 nonreinforced trials and the mean of his largest response to any one of the four CS− lights on each trial (MCR−)

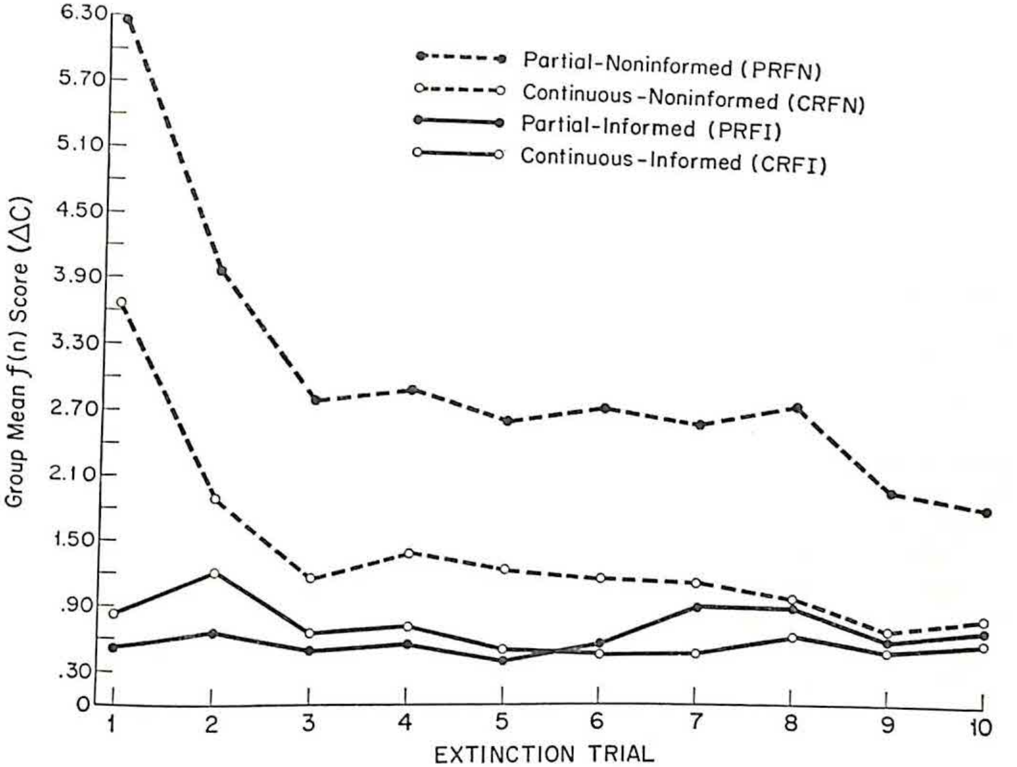


FIG. 1. Group mean responses on the first 10 extinction trials.

for these same 15 nonreinforced trials. Assuming that identical results would exist for the CRF groups, could their MCR+ levels be obtained, the MCR− score for each *S* in the four groups was used as the best estimate of $R(1)$. Each *S* in these four groups was assigned an $f(n)$ score on the first 10 extinction trials based on his own acquisition MCR−. This transformed score equates *S*s in terms of differences in nonspecific GSR lability. The mean $f(n)$ score in micromhos for each of the four groups on the first 10 extinction trials is presented in Fig. 1.

Table 1 presents the results of trend analyses (Edwards, 1960) performed on these data. In the Noninformed groups, the significant Groups effect, $F(1, 38) = 6.01$, $p < .025$, demonstrates the classical PRE with the PRFN group showing greater resistance to extinction than the CRFN. In addition, using the conservative tests suggested by Greenhouse and Geisser (1959), significant Trials, $F(1, 38) = 7.45$, $p < .01$, and Groups × Trials interaction effects, $F(1, 38) = 7.84$, $p < .01$, were found indicating that a response decrement oc-

TABLE 1

ANALYSIS OF VARIANCE OF THE FIRST 10 EXTINCTION TRIALS

Source	*df*	Noninformed Groups		Informed Groups	
		MS	*F*	*MS*	*F*
Groups (G)	1	278.43	6.01*	.08	.03
Error (a)	38	46.29		3.19	
Trials (T)	9	43.46	7.45**	.77	2.66
G × T	9	45.70	7.84**	1.88	6.48*
Error (b)	342	5.83		.29	

* $p < .05$.
** $p < .01$.

TABLE 2

Analysis of Variance of the First 5 Extinction Trials: Informed Groups

Source	df	MS	F
Groups (G)	1	.82	.53
Error (a)	38	1.54	
Trials (T)	4	5.24	10.92**
G × T	4	1.63	3.39
Error (b)	152	.48	

** $p < .01$.

curred and that the groups differed in their rates of extinction. In the Informed groups, the lack of a significant Groups effect, $F\ (1, 38) < 1$, is interpreted as demonstrating the abolition of the PRE. The lack of a significant Trials effect, $F\ (1, 38) = 2.66$, $p > .05$, as well as the significant interaction effect, $F\ (1, 38) = 6.48$, $p < .05$, may have been due to the reversal of the mean levels of the two groups over the last five trials as compared to the first five. Therefore, the data of only the first five trials were subjected to an additional analysis and the results are presented in Table 2. The Groups effect remained nonsignificant, $F\ (1, 38) < 1$. A significant Trials effect was present, $F\ (1, 38) = 10.92$, $p < .01$, indicating an overall response decrement across trials. Combined with the lack of a significant interaction, $F\ (1, 38) = 3.39$, $p > .05$, the results indicate the decrement occurred equally in both groups across the first five trials.

Pseudoconditioning.—One of the most effective methods of controlling for pseudoconditioning is to compare each *S*'s response to the CS (CR+) with his response to a nonreinforced stimulus presented within a short time interval (CR−). During extinction the pseudoconditioning response level may also be considered as the extinction level of the CR+. Therefore, each *S*'s largest response to any of the four nonreinforced lights during extinction was deemed to be a pseudoconditioning response, and t tests were performed on each group's transformed extinction MCR+ and MCR− scores. The results of these analyses are presented in Table 3. The results demonstrate that, for all groups, the MCR+ over the first five trials was significantly greater than the MCR−. On the second five trials the two Noninformed groups' MCR+ scores were still significantly higher than their MCR− scores (PRFN: $p < .05$; CRFN: $p < .01$). For the two Informed groups, however, there was no significant difference ($p > .05$) between the MCR+ and the MCR− scores and extinction was considered to have occurred.

TABLE 3

Extinction Pseudoconditioning Analyses

Group	First Five Trials			Second Five Trials		
	Mean CR+	Mean CR−	t	Mean CR+	Mean CR−	t
PRFI	.522	.116	3.87**	.739	.520	1.76
CRFI	.785	.194	4.15***	.533	.695	.65
PRFN	3.714	.347	3.77**	2.423	.882	2.71*
CRFN	1.836	.304	9.86***	.962	.678	3.37**

Note.—$df = 19$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Extinction criterion.—This method utilized a nonparametric analysis of the data. A conditioned response (CR) was defined as the GSR evoked by the CS on each trial which was greater in magnitude (ΔC) than the GSR evoked by any of the four neutral stimuli. Each *S* was then assigned a plus or minus on each of the 30 extinction trials depending upon whether or not he met the criterion for a CR. Extinction was considered to have occurred when a CR failed to be elicited on three successive trials. Each *S* was assigned a score consisting of the number of trials before the extinction criterion was met, with a maximum score of 30. A one-tailed Mann-Whitney *U* test was used to analyze the results. The PRE was present in the Noninformed groups with $U = 106.5$, $p < .01$ and PRFN > CRFN. For the Informed groups no statistically significant difference could be demonstrated with $U = 183.5$, $p > .05$ and CRFI > PRFI. This result is interpreted as demonstrating the abolition of the PRE in the Informed groups. The results of this method are in agreement with those of the shape-function method in that the PRE was present in the Noninformed and had been abolished in the Informed groups.

Discussion

The present results support the hypothesis that when ease of discrimination between an acquisition and extinction series is equalized for a PRF and CRF group, the PRE will no longer be present. However, the presence of the PRE in the Noninformed group and its absence in the Informed suggests that the CR has two separate components. These components may be categorized as a simple vs. a mediated CS-UCS relationship. The following discussion of these components is based in part on articles by Mowrer (1938) and Razran (1961).

In the simple CR the CS directly signals, or is a surrogate for, the UCS. In the mediated CR the CS signals an anticipation (e.g., fear) of the reinforcement. This anticipation occurs when the CS is part of a compound stimulus including both environmental contingency cues (e.g., shock electrodes) and pattern of reinforcement. A given CR may contain both simple and mediated components. In the present experiment the PRF and CRF groups acquired different configurational CRs as a function of the patterns of reinforcement. The extinction instructions and removal of the environmental contingency cues eliminated the mediated CR for the Informed groups, leaving only the simple CR to be extinguished. The Noninformed groups, however, had to extinguish not only the simple CR but also the mediated CR. Dependent as it is upon the stimulus compound of both contingency cues and pattern of reinforcement, the mediated CR was responsible for the classical PRE. The methodology of the present experiment allows the separation of these two components of the CR and demonstrates that they obey different laws in regard to resistance to extinction.

In keeping with the results of most experiments in this area, the instructions facilitated the extinction of the conditioned GSR. However, contrary to the results obtained by Wickens, Allen, and Hill (1963), a significant difference was found between the Informed and Noninformed groups on the very first extinction trial, PRF: $t\ (38) = 2.50$, $p < .05$ with PRFN > PRFI and CRF: $t\ (38) = 4.10$, $p < .001$ with CRFN > CRFI. Using a postexperimental inquiry, Wickens et al. (1963) reported, "Slightly less than one in five *S*s reported that they were not confident that the shock would not be presented [p. 236]." However, the response protocols of these nonconfident *S*s were included in the comparison of the Informed and Noninformed groups. The present experiment revealed a clear difference between the Informed and Noninformed groups on the first trial by excluding the protocols of the nonconfident *S*s.

The present results are relevant to several other issues. Grant and Schipper (1952) have presented a two-factor theory of the PRE. These investigators postulated that resistance to extinction is directly related to the number of reinforcements (or response strength) and inversely related to the discriminability of the acquisition from the extinction series. The Grant and Schipper hypothesis would predict that when discrimination has been equalized, the CRF group would show greater resistance to extinction than the PRF group. The present results do not uphold this deduction. However, in terms of both the shape-function analysis of the first 5 extinction trials and the nonparametric analysis of the 30 trials, there was a statistically nonsignificant tendency for the CRFI group to show greater resistance to extinction than the PRFI group. Whether the failure to uphold the Grant and Schipper hypothesis is due to the strong UCS at a 500-msec. ISI producing a conditioning asymptote, or whether, in aversive conditioning, a nonreinforced trial might be incremental to habit strength, remains open for further investigation.

Finally, Chatterjee and Eriksen (1962) have maintained that cognitive expectancy is an essential correlate of heart-rate responses. In the present experiment the Informed groups demonstrated a conditioned GSR contrary to their cognitive expectancies.

REFERENCES

Anderson, N. H. Comparison of different populations: Resistance to extinction and transfer. *Psychol. Rev.*, 1963, **70**, 162–179.

Bitterman, M. E., Feddersen, W. E., & Tyler, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

Bridger, W. H., & Mandel, I. J. A comparison of GSR fear responses produced by threat and electric shock. *J. psychiat. Res.*, 1964, **2**, 31–40.

Chatterjee, B., & Eriksen, C. W. Cognitive factors in heart rate conditioning. *J. exp. Psychol.*, 1962, **64**, 272–279.

Edwards, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Holt, Rinehart, & Winston, 1960.

Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–520.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. *Psychometrika*, 1959, **24**, 95–112.

Jenkins, H. M. Resistance to extinction when partial reinforcement is followed by regular reinforcement. *J. exp. Psychol.*, 1962, **64**, 441–450.

Mowrer, O. H. Preparatory set (expectancy): A determinant in motivation and learning. *Psychol. Rev.*, 1938, **45**, 62–91.

Mowrer, O. H., & Jones, H. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 293–311.

O'Connell, D. N., & Tursky, B. Silver-silver chloride sponge electrodes for skin potential recording. *Amer. J. Psychol.*, 1960, **37**, 302–304.

O'Connell, D. N., Tursky, B., & Orne, M. T. Electrodes for the recording of skin potential: An evaluation. *Arch. gen. Psychiat.*, 1960, **3**, 252–258.

Razran, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81–147.

Theios, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 1–6.

Weinstock, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

Wickens, D. D., Allen, C. K., & Hill, F. A. Effect of instructions and UCS strength on extinction of the conditioned GSR. *J. exp. Psychol.*, 1963, **66**, 236–240.

Wilson, W., Weiss, E. J., & Amsel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1955, **50**, 51–60.

(Received February 27, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 483–489

STRENGTH OF THE RELATIONSHIP BETWEEN THE VALUE OF AN EVENT AND ITS SUBJECTIVE PROBABILITY AS A FUNCTION OF METHOD OF MEASUREMENT [1]

DEAN G. PRUITT AND ROBERT D. HOGE

Center for Research on Social Behavior, University of Delaware

This study was designed to evaluate the strength of the relationship between the value of an event and the subjective probability of that event as measured by 3 distinct techniques. A positive relationship was found between these 2 variables under all conditions but diminished in strength as pressures for accuracy increased. The results were discussed in the light of Rotter's criticism of earlier studies. In addition, a comparison was made of the 3 methods of measurement, from which probability ratings emerged as the most reliable technique for studying reactions to objective probability displays.

Several studies have been published that show a positive relationship between the value of an event and the likelihood that a person will predict its occurrence (Crandall, Solomon, & Kellaway, 1955; Irwin, 1953; Marks, 1951). The results of these studies have sometimes been cited as evidence that the value of an event can have an effect on its subjective probability (SP). However, Rotter (1954) has registered an objection to this interpretation, arguing that the value of the events in these studies may not have influenced the SP of their occurrence but only "the potential of the subject's stating that he thinks they will occur, partly because he has no need to differentiate his true expectancies (SP) from his wishes [p. 164]." Considerable theoretical importance attaches to the issue of whether the value and SP of an event are interrelated, because the dominant model of risky decision making, the subjectively expected utility (SEU) model, makes no provision for such an interrelationship (Edwards, 1962).

[1] This research was supported by the Air Force Office of Scientific Research under Contract No. AF 33(616)-7621 and by the National Science Foundation under Research Grant No. GS-311.

The present study had two main goals:

1. To replicate the major features of the earlier experiments using three methods for measuring SP: a guessing, a probability rating, and a betting method. We reasoned that it would be harder to maintain Rotter's position that SP is not involved in this kind of study if results similar to those found in earlier experiments were obtained under all three methods of measuring SP.

2. To test the hypothesis implied by Rotter that predictions concerning an event will be less affected by the value of that event the stronger the pressures for accuracy. Assuming that the rating method would involve greater pressure for accuracy than the guessing method, we predicted that the value of an event would have more effect on the measured SP[2] of that event under the latter method. In

[2] Throughout this report a distinction is made between SP and measured SP. The former is a psychological process; the latter an experimental observation based on one or another technique of measurement. This distinction must be made because Rotter's criticism amounts to the assertion that differences in measured SP do not reflect differences in true SP.

addition, we introduced a fourth condition in which incentives for accuracy were coupled with the rating method. We predicted that the measured SP of an event would be less strongly influenced by the value of that event under this condition than under the simple rating condition. No predictions were made about the betting method, since we were not sure beforehand how much pressure for accuracy it would exert.

Method

Apparatus and procedure.—The apparatus consisted of two 40-w. light bulbs mounted side by side behind a translucent screen. The points on the screen where the lights could be seen were labeled "A" and "B."

During each trial, 24 flashes of light were shown, each flash coming from either Light A or Light B. Both the duration of a flash and the interval between flashes were approximately 1 sec. long. The sequence in which the lights were flashed was controlled from a programing apparatus.

The *S* sat at a distance of about 8 ft. from the screen. In an initial briefing, he was told that the sequence of lights was controlled by a chance device and that the probability of Light A (and therefore of Light B) changed from trial to trial. He was also led to believe that every sequence terminated in a twenty-fifth flash of light, which would be recorded for *E*'s use but would be obscured from *S*'s view.

On each trial, after viewing the first 24 flashes, *S* was asked to make a judgment about the twenty-fifth flash. The form of the judgment depended on the condition to which he had been assigned:

1. In the *guessing condition*, *S* was simply instructed to circle one of two letters, "A" or "B," depending on which light he thought would appear on the twenty-fifth flash.
2. In the *rating condition*, *S* was asked to circle a probability level on the following scale:

Probability of Light "A" appearing

1.0	.9	.8	.7	.6	.5	.4	.3	.2	.1	0
"A" more likely than "B"						"B" more likely than "A"				

3. The same scale was used in the *rating with incentive condition*, but here the instructions stressed accuracy, and *S* was informed that a cash reward would be given to the two *S*s whose estimates of the true probabilities were most "accurate."
4. In the *betting condition*, the following statement appeared at the bottom of each answer sheet: "For the bet 'If light A win 100 points' I am willing to pay up to ———— points."

The *S*s were told that they were playing a betting game, whose currency consisted of points rather than money. On each trial they had an opportunity to bid for the wager, "If A win 100 points," which means that if Light A comes on you get 100 points and if Light B comes on you get nothing. Bids were made by writing a number between 0 and 100 in the space provided. The *S*s were urged to bid the maximum amount they were willing to pay for the bet, on the assumption that an independent determination of the worth of the bet would be made and that they would have to purchase the bet, at the amount of the independent determination, if their bid was equal to or greater than the independent determination. They were told that the bets would be played off after the session and that the two persons with the highest points would be given money prizes. They were urged to vary the amount bid in terms of the likelihood, as they saw it, that Light A would come on during the twenty-fifth trial.

The judgments were recorded in a booklet containing 1 page for each trial. Each page was divided into two parts. In the top part was a statement about the amount to be credited or debited to *S* if Light A occurred on the twenty-fifth flash. There were five possible statements: "If light A win 50 cents," "If light A win 10 cents," "If light A win or lose nothing," "If light A lose 10 cents," "If light A lose 50 cents." In the bottom part of the page was a place for *S* to record his response.

In all conditions, *S*s were run in groups that varied in size from two to five. The *E* sat in the room with *S*s, operating the program by remote control. Each session lasted approximately 1 hr. At the beginning of a session, *S*s were credited with $1.25 and in-

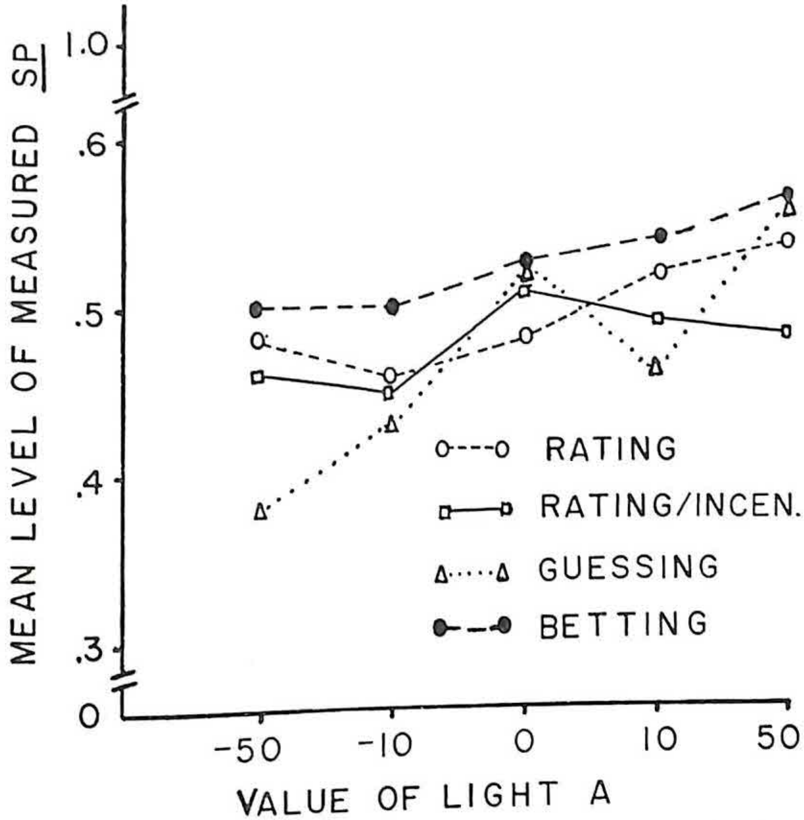


FIG. 1. Average levels of measured SP for the five levels of value.

formed that their losses would come out of this initial payment. At the end of a session, wins and losses were calculated and payoffs made.

Design and subjects.—A factorial design was employed, involving four variables: (*a*) the four *conditions* described above; (*b*) three basic *sequences* of lights: one in which Light A appeared 33% of the time, one in which it appeared 50%, and one in which it appeared 67% of the time;[3] (*c*) the five levels of *value* associated with the occurrence of Light A: win 50¢, win 10¢, win or lose nothing, lose 10¢, and lose 50¢; and (*d*) two *replications* of the design.

Eighty volunteers from an undergraduate course in psychology served as *Ss*, 20 being assigned at random to each of the four conditions. Each *S* participated in 30 trials, composed in the following way: the three basic sequences of lights were exhaustively combined with the five values, and each of these combinations was presented twice. The order of trials varied from *S* to *S* on a random basis.

[3] The sequences of lights were as follows, 33%: BAABBBBABBBABABBABBBABBA, 50%: AAABBAABBABBABABAAABBABB, 67%: AABABAAABAABABAABBAAABAA.

RESULTS

In the guessing method, tests of significance were performed on binary data, a guess of Light A being given a score of "1" and a guess of Light B a score of "0." Group measures of SP were obtained in this condition by calculating the percentage of all guesses in a group that named Light A. In the rating conditions, the numbers circled by *S* were used as indexes of SP in all calculations. In the betting condition, SP was calculated for each trial by dividing *S*'s bid by 100, a standard technique (Preston & Baratta, 1948; Shuford, 1959).

TABLE 1

Analyses of Variance for the Four Conditions

Source	df	F			
		Guessing	Rating	Rating with Incentive	Betting
Subjects (A)	19	1.72*	2.44**	1.11	7.11***
Values (B)	4	3.87**	2.74*	2.22	1.95
Linear trend	1	12.60***	6.43**	2.59	5.57*
Sequences (C)	2	17.26***	35.44***	35.19***	11.14***
A × B	76	1.05	1.52*	.85	1.01
A × C	38	4.13***	7.90***	6.84***	6.71***
B × C	8	.76	1.38	.87	2.56*
A × B × C	152	1.02	.82	.75	1.02
Replications	300	(0.168)	(2.92)	(3.57)	(382.94)
Total	599				

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Relationship between value and measured SP.—Evidence of a positive relationship between value and measured SP can be seen in all four curves of Fig. 1. The statistical significance of this evidence is given by the "linear trend over values" terms in the analyses of variance presented in Table 1. This term reaches an acceptable level of significance in the case of the guessing, rating, and betting conditions and approaches significance in the case of the rating with incentive condition.

A visual comparison of the slopes of the curves in Fig. 1 suggests, as predicted, that the relationship between value and measured SP is more pronounced under the guessing than under the rating condition. In order to make an adequate test of this comparison, it was necessary to reduce the data from these two conditions to the same scale. This was done by dichotomizing the rating data, treating ratings between .6 and 1.0 as guesses of Light A and ratings between 0 and .4 as guesses of Light B. Although ratings of .5 actually represented a failure to guess, it was necessary to use them to obtain a set of data that would be comparable to the guessing data. They were arbitrarily treated as guesses of Light B, on the assumption that the constant error which this procedure would introduce into the averages would have no effect on the slope of the curve relating the percentage of guesses to value. Using a technique for comparing linear trends described by Winer (1962, p. 276), a comparison was made between the slope of the curve for guessing data shown in Fig. 1 and the slope of a comparable curve for the dichotomized rating data. A t was obtained of 1.85, which was significant beyond the .05 level with a one-tailed test of significance.

Inspection of Fig. 1 also supports the prediction that the relationship between value and measured SP would be stronger under the rating than under the rating with incentive condition. However, with a t of .93, this can only be considered a trend in the predicted direction.

Individual differences in the degree to which measured SP is affected by value are reflected in the *S*s × Values interaction, which reaches significance only in the rating method. In a

side analysis, a correlation ($r = .34$, $p = .06$) was found between this effect and scores on the *Pt* scale of the MMPI, suggesting the existence of a direct relationship between psychasthenia and susceptibility to value distortions of SP, as measured by the rating method.

Sequence effects.—Average measured SPs for the three basic sequences of lights are shown in Table 2. As might be expected, the averages for the sequence containing 33% of Light A are close to .33. But the averages for the other two sequences seem totally unrelated to the frequency of Light A; the averages for the 50% sequence are much too high, and for the 67% sequence much too low. Table 1 shows that the variance due to sequence differences is highly significant for all four methods, indicating that sequence differences had an important, if not predictable, effect on measured SP.

Table 1 also reveals a highly significant *Ss* × Sequences term for every method. This implies that people differ markedly and consistently in the techniques they use for deriving the SP of an event from sequential information.

Comparison of the methods of measurement.—Because identical experiments had been done under all four conditions, it was possible to compare the methods of measurement with regard to the kind of information transmitted. An analysis of variance was performed on each of the 80 *Ss*, dividing the variance between his responses into four components: (*a*) that attributable to differences between the three sequences of lights, (*b*) that attributable to differences between the five outcome values, (*c*) that attributable to the interaction between sequences and values, and (*d*) that attributable to differences between the two replications of the experimental conditions. The four variance numbers (*SS*s) for each *S* were then converted into *proportions of the total variance* for that *S*. This produced a matrix of 320 proportions, composed of four columns (corresponding to the components of variance) and four groups of 20 rows (corresponding to the four conditions and the 20 *Ss* within each condition), with the proportions in every row summing to 1. Averages, computed across the 20 *Ss* in each group, produced the data shown in Table 3.

TABLE 2

MEAN MEASURED SP FOR THE THREE DIFFERENT SEQUENCES OF LIGHT

	Sequence		
	33% A	50% A	67% A
Guessing	.29	.75	.38
Rating	.32	.72	.45
Rating with incentive	.28	.69	.46
Betting	.41	.65	.51

The methods clearly have different profiles. We can distinguish two groups: one consisting of the rating and rating with incentive methods, the other consisting of the guessing and betting methods. These groups differ primarily in the proportion of variance associated with sequences: differences between the sequences of lights having more effect on behavior for the first group of methods than for the second. In addition, the second group had a larger proportion of variance associated with replications and a somewhat larger proportion with values. It was possible to test the significance of the difference between methods by performing a second analysis of variance on the 320 proportions described just above. The interaction between methods of measurement and

TABLE 3

Average Proportions of Variance Attributable to Each of Four Components of Variation

Cond.	Components of Variation			
	Values	Sequences	Val. × Seq.	Replications
Rating	.0918	.6012	.1007	.2058
Rating with inc.	.0638	.5913	.0960	.2489
Guessing	.1142	.3496	.1860	.3502
Betting	.1037	.3706	.1725	.3528

components of variation was highly significant, $F(9, 228) = 7.38, p < .001$.

Discussion

The results provide new evidence about the relationship between the value of an event and predictions concerning that event but, unfortunately, do not enable us, as had been hoped, to accept or reject the hypothesis that value affects true SP. On the one hand, the emergence of a positive relationship between value and measured SP under three rather different methods of measurement might be thought to strengthen the contention that value affects true SP. On the other hand, the diminution of this relationship with increased pressure for accuracy seems to provide support for Rotter's position. Actually, the implications of this latter evidence are themselves quite unclear. We might conclude, with Rotter, that pressures for accuracy affect the process through which SP is translated into behavior, causing people to filter out other forces such as values and pay closer attention to their SPs. Or we might equally logically conclude from these findings that, under pressures for accuracy, people are stricter about the sources of their SP. Further research is needed to clarify this ambiguity.

For all four conditions, a marked discrepancy was found between the proportion of times Light A was exhibited in the 50% and 67% sequences and the measured SP of Light A. These discrepancies may have been due to the negative recency effect or "gambler's fallacy," which makes people more likely to predict one kind of event the longer the run of the other kind of event that has just preceded (Jarvik, 1951). The 50% sequence ended in two flashes of Light B, which would lead to the perceived overestimation of the probability of Light A if the negative recency effect were in operation. The 67% sequence ended in two flashes of Light A, which may help account for the underestimation of SP at the end of that sequence. The 33% sequence ended in a single flash of Light A, which would not be expected to produce much distortion.

In comparing the methods of measurement, we found that people were more influenced by the sequences of lights in the rating methods than in the guessing or betting methods. The sequences of lights are, of course, the only evidence which is objectively relevant to probability—the only information on which it is reasonable to base SP. We might, therefore, conclude that people are more influenced by objective evidence when making probability ratings than when guessing or betting. They may use this evidence wrongly and commit the gambler's fallacy, but they are still more likely to use it. If this is true, it may make methodological sense to use rating rather than guessing or betting methods for studying reactions to objective probability displays.

REFERENCES

Crandall, V. J., Solomon, D., & Kellaway, R. Expectancy statements and decision times as functions of objective probabilities and reinforcement values. *J. Pers.*, 1955, **24**, 192–203.

Edwards, W. Utility, subjective probability, their interaction, and variance preferences. *J. conflict Resolut.*, 1962, **6**, 42–51.

Irwin, F. W. Stated expectations as functions of probability and desirability of outcomes. *J. Pers.*, 1953, **21**, 329–335.

Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

Marks, R. The effect of probability, desirability and 'privilege' on the stated expectations of children. *J. Pers.*, 1951, **19**, 332–351.

Preston, M. G., & Baratta, P. An experimental study of the auction value of an uncertain outcome. *Amer. J. Psychol.*, 1948, **61**, 183–193.

Rotter, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

Shuford, E. H. A comparison of subjective probabilities for elementary and compound events. Report No. 20, 1959, University of North Carolina, Psychometric Laboratory.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received February 27, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 490–495

EFFECT OF PAIRING DIRECTIONALITY AND ANTICIPATORY CUE IN PAIRED-ASSOCIATE LEARNING [1]

JAMES F. VOSS

University of Pittsburgh

Each of 4 groups was presented a list of 8 paired-associate trigrams for 20 trials. A 2 × 2 factorial design was employed in which stimulus-cue variation and pairing direction were varied. The 4 experimental conditions thus were: (a) A, A-B; (b) A, A-B; B, A-B; (c) A, A-B; A, B-A; (d) A, A-B; B, A-B; A, B-A; B, B-A. It was found that cue variation yielded significantly fewer correct responses. Further analyses of response learning, associative learning, correct response/opportunity ratio, and errors suggested that this decrement was due to the necessity of learning both items of each pair as responses. Pairing direction variation had a smaller, yet significant effect which was related to intrusions. No interactions were found.

The paired-associate (PA) anticipation method consists of presentation of a cue item (A) followed by the pairing of the cue item and another item (A-B).[2] Experimental results (Feldman & Underwood, 1957; Jantz & Underwood, 1958; Murdock, 1956, 1958; Richardson, 1960) have demonstrated that not only A-B associative strength develops in PA anticipation learning, but backward associative strength, i.e., B-A, also develops.

The issue of bidirectional associative strength has been pursued further by Asch and Ebenholtz (1962), Underwood and Keppel (1963), and Schilde (1963). Asch and Ebenholtz, using the anticipation and recall methods of PA learning, interpreted their results as indicating that associations are of equal bidirectional associative strength. Underwood and Keppel (1963) compared acquisition of a standard anticipation PA list to an experimental list in which the cue item and related pairing occurred as A, A-B on one-half of the trials and B, B-A on the other one-half of the trials. The results were interpreted as indicating that "bidirectional learning is a simple summation of time to learn forward and backward associations to the same level of strength [p. 474]." Using anticipation and recall procedures and a design highly similar to Underwood and Keppel, Schilde found significantly more errors in the bidirectional condition.

Consideration of the Underwood and Keppel paradigm indicates that the experimental list involved concomitant variation of the item that was used as the cue (A or B) with the direction of the pairing (A-B or B-A, respectively). The purpose of the present experiment was to determine the independent and joint effects of the cue item and direction of pairing variables.

[1] This research was supported by the National Institute of Mental Health (MH-08291-01) and Child Health and Human Development (HD-00957-01). The author wishes to thank Patricia Polesiak for assistance in the gathering and analyzing of the data.

[2] Typically, forward learning has been termed S-R and backward learning R-S. However, the A-B nomenclature is employed in this report in order to avoid confusing statements such as "The R item is the stimulus – – –."

TABLE 1

PERFORMANCE MEASURES OF THE EXPERIMENTAL CONDITIONS

Cond.	Column[a]									
	1	2	3	4	5	6	7	8	9	10
C_1D_1	52.7	8.5	—	1.9	64.2	—	40.2	.57	19.3	2.7
C_1D_2	48.9	8.6	—	1.5	65.1	—	38.8	.51	31.4	9.5
C_2D_1	39.7	8.6	15.9	2.5	64.6	92.1	34.3	.43	17.9	3.5
C_2C_2	25.4	10.1	16.4	3.3	60.5	75.0	24.7	.36	20.1	7.6

[a] The columns denote the following measures: 1, mean correct responses; 2, mean trials to the first response; 3, mean trials to the first occurrence of the second response of each pair; 4, mean associative learning trials; 5, percent correct responses on the first response occurrence; 6, percent correct responses on the first occurrence of the second response item of each pair; 7, percent correct responses on Trial $n+1$; 8, mean correct response/opportunity ratio; 9, percent intralist intrusions; 10, percent extralist intrusions.

METHOD

Procedure.—Each *S* was presented eight paired associates by the anticipation method for 20 trials. A different random order of the paired associates was used on each trial. A 2:2-sec. presentation rate was employed for the cue and pairing presentations with a 4-sec. intertrial interval. Standard paired-associate anticipation instructions were used.

Anticipatory cue item (C) and directionality of pairing (D) were varied in a 2×2 factorial design. Condition C_1D_1 involved presentation of the same cue item and the same direction of pairing (A, A-B). Condition C_1D_2 was identical to Cond. C_1D_1 except that the pairings of each paired associate were reversed on one-half of the trials (A, A-B; A, B-A). A different random order of the A-B and B-A pairings was employed for each association. Condition C_2D_1 was identical to Cond. C_1D_1 except that Item A of each association was employed as the cue on one-half of the trials and Item B on the other one-half of the trials (A, A-B; B, A-B). A different random order was employed for each association with respect to A and B trials. Condition C_2D_2 was identical to Cond. C_1D_1 except that the cue item was varied on the same trials as Cond. C_2D_1 and the pairing direction was varied on the same trials as Cond. C_1D_2 (A, A-B; B, A-B; A, B-A; B, B-A). The random order of the cue and pairing trials of each association was restricted so that the cue and pairing direction variations were orthogonal. Thus, in Cond. C_2D_2, each combination of cue and pairing direction was presented for 5 of the 20 trials for each association.

Materials.—The A-B pairs were 47–53% Glaze association value trigrams which were scaled by Battig (1959) such that A-B associative strength is of approximately the same magnitude as B-A associative strength. The trigrams were pronounced by *Ss* with articulation occurring only as anticipations.

Subjects.—The *Ss* were 60 introductory psychology students at the University of Pittsburgh and they were assigned to each of the four conditions ($N = 15$/condition) by use of a table of random numbers.

RESULTS AND DISCUSSION

Total correct responses.—The first column of Table 1 presents the mean correct responses for the experimental conditions. An analysis of variance performed on the correct response frequency data revealed a significant C source of variation, $F\ (1, 56) = 16.59$, $p < .01$, and a significant D source of variation, $F\ (1, 56) = 4.10$, $p < .05$. The $C \times D$ interaction source of variation is not significant. Thus, the findings indicate that use of only one item of each association as cue yielded considerably better performance than use of both items as cues and undirectional pairing presentation yielded better performance than bidirectional pairing presentation.

Response learning.—Response learning was measured by tabulating the trials to *S*'s first response occurrence of each association, regardless of whether the response was correct or incorrect (Postman, 1962). Before pre-

senting the data, however, a methodological issue must be considered.

When A and B were employed as cues, response learning was required for both items. Two tabulations therefore were performed for Cond. C_2D_1 and C_2D_2. The first was trials to the occurrence of the first response, regardless of whether it was A *or* B; the second was trials to the first occurrence of both items as responses, A *and* B.

Column 2 of Table 1 presents mean trials to the first response for the four experimental conditions. Any response not given was designated Trial 21. An analysis of variance performed upon the trials to the first response data revealed that the C, D, and C $\times$ D sources of variation are not significant.

Column 3 of Table 1 presents mean trials to the first occurrence of A *and* B as responses. Any response not given again was designated Trial 21. It should be noted that the mean trials from first to second response learning was 7.3 for Cond. C_2D_1 and 6.3 for Cond. C_2D_2. An analysis of variance performed upon the C_2D_1 and C_2D_2 second response data combined with the C_1D_1 and C_1D_2 data revealed a significant C source of variation, $F\ (1, 56) = 45.05$, $p < .01$. The D and C $\times$ D sources of variation are not significant. It should be pointed out that since either A or B was presented as cue between the occurrence of the first and second responses in Cond. C_2D_1 and C_2D_2, the opportunity for the second item to occur as a response was a function of the trials on which it was a response rather than a cue. The data indicate that of the 7.3 trials in Cond. C_2D_1 between the first- and second-item response learning, a mean of 3.8 trials occurred in which the response was the second response item of the 6.3 trials in Cond. C_2D_2 between the first- and second-item response learning, a mean of 3.5 trials occurred in which the response was the second response item.

In summary, response learning does not vary significantly as a function of cue variation when the first response in Cond. C_2D_1 and C_2D_2 is considered. However, response learning takes significantly more trials in Cond. C_2D_1 and C_2D_2 than C_1D_1 and C_1D_2 when both items of each pair are considered. In addition, pairing directionality does not significantly influence response learning.

Associative learning.—Associative learning was measured by tabulating the following difference: trials to the first correct response minus trials to the first occurrence of that particular response (response learning) for each pair (Underwood & Schulz, 1960). Column 4 of Table 1 presents the mean associative learning trials for each experimental condition. The means of Cond. C_2D_1 and C_2D_2 refer to the first item of each pair correctly associated. (This item was almost without exception the first item learned as a response.)

The mean associative learning trials for each *S* was computed and the data were submitted to an analysis of variance. The C source of variation is significant, $F\ (1, 56) = 8.40$, $p < .01$. The D and C $\times$ D sources of variation are not significant.

Since the opportunity for correct association was less in Cond. C_2D_1 and C_2D_2 than in Cond. C_1D_1 and C_1D_2 because of the cue variation, the trials for each association were tabulated on which the item given previously as a response could be given as a correct association. The data indicate that the mean associative learning trials, taking into account opportunity for a correct association, were 1.3 in Cond. C_2D_1 and 1.3 in Cond. C_2D_2. These data were con-

sidered with the C_1D_1 and C_1D_2 data and submitted to an analysis of variance. The C, D, and C × D sources of variation are not significant. These results indicate that the significant effect of cue variation upon associative learning is attributable to the lack of opportunity that cue variation provides for a correct association to be made.

Associative learning trials were tabulated for the second item of each pair in Cond. C_2D_1 and C_2D_2. The mean associative learning trials were 0.3 in Cond. C_2D_1 and 1.2 in Cond. C_2D_2. These data indicate that associative learning of the second response item was more rapid than that of the first item. An analysis of variance was not performed on these data because of the restriction produced by the relatively large frequency of responses that were never associated correctly (47 of 120 in Cond. C_2D_1 and 58 of 120 in Cond. C_2D_2).

Associative learning and its relation to response learning were studied further by consideration of the following question: what percentage of first response occurrences was correct responses? Column 5 of Table 1 presents the percent correct responses given on the first occurrence of the response. The percent of Cond. C_2D_1 and C_2D_2 represents the first response item learned. The percent correct responses on the first response occurrence was computed for each *S*, submitted to an arcsin transformation, and the transformed data were submitted to an analysis of variance. The C, D, and C × D sources of variation are not significant.

Column 6 of Table 1 presents the percent correct responses that were given on the first occurrence of the second response item in Cond. C_2D_1 and C_2D_2. The data were not submitted to an analysis of variance because of the relatively large number of *S*s that gave all correct responses on the first occurrence of that particular response (7 of 15 in Cond. C_2D_1 and 6 of 15 in Cond. C_2D_2).

In summary, the associative learning data indicate: (*a*) associative learning of the first item in Cond. C_2D_1 and C_2D_2 is poorer than C_1D_1 and C_1D_2, but this difference is due primarily to the lack of opportunity in the former conditions to associate the first response item correctly; (*b*) associative learning of the second response item in Cond. C_2D_1 and C_2D_2 was quite rapid; (*c*) pairing direction yielded no significant effect upon associative learning.

Post-correct-response trial.—The trial on which a response is first given correctly is denoted Trial n and the subsequent trial, $n + 1$. Column 7 of Table 1 presents the percent correct responses on Trial $n + 1$ with the first response item considered in Cond. C_2D_1 and C_2D_2. The percent correct on Trial $n + 1$ was tabulated for each *S*, submitted to an arcsin transformation, and the transformed data were submitted to an analysis of variance. Despite the relatively large group differences, the C, D, and C × D sources of variation are not significant. It should be noted that this method of tabulation weights each *S* equally, but the number of $n + 1$ trials within a given *S* is not differentially considered.

The percent correct responses on Trial $n + 1$ in Cond. C_2D_1 and C_2D_2 was tabulated when the same cue item was presented on Trials n and $n + 1$ and when the Trial $n + 1$ cue item was different from the Trial n cue item. In Cond. C_2D_1, the percent correct responses for the same cue item was 43.2 and for the second cue item was 26.6; in Cond. C_2D_2, the percent correct responses for the same cue item was 35.7 and for the second cue item was 18.9. The percent correct on Trial $n + 1$ for each *S* under the

same and second cue conditions was submitted to an arcsin transformation and a t test was performed on the transformed data. In Cond. C_2D_1, no significant difference was found between percent correct on $n + 1$ of the same cue and second cue occurrence, t (26) = 1.71, $p > .05$.[3] However, significantly greater percent correct responses occurred in Cond. C_2D_2 when the same cue item occurred on Trials n and $n + 1$, t (26) = 1.84, $p < .05$.

The $n + 1$ data thus indicate that: (*a*) use of both items of a pair as cues yielded a tendency toward fewer correct responses on Trial $n + 1$, especially when the cue item changed over Trials n and $n + 1$; (*b*) pairing direction had no significant effect.

Correct opportunity ratio.—The ratio of correct responses following the first correct response + opportunity for correct responses after the first correct response was tabulated for each pair. The mean ratio for each experimental condition is presented in Column 8 of Table 1. The mean ratio for each S was submitted to an arcsin transformation and the transformed data were submitted to an analysis of variance. The C source of variation is significant, F (1, 56) = 11.92, $p < .01$. The D and C $\times$ D sources of variation are not significant. These results indicate that cue variation yielded a significantly detrimental effect upon correct responses when the opportunity for correct responses is taken into account.

In order to isolate this detrimental effect, the correct/opportunity ratio was tabulated for each of the two response items in Cond. C_2D_1 and C_2D_2. In Cond. C_2D_1, the mean C/O ratios were .54 and .63 for the first and second response items, respectively; in Cond. C_2D_2, the mean C/O ratios were .49 and .65 for the first and second response items, respectively.

These ratios indicate that first, the C/O ratio for the four experimental conditions is quite comparable when the first response item is considered in Cond. C_2D_1 and C_2D_2. Second, the C/O ratio for the second response item in Cond. C_2D_1 and C_2D_2 is greater than that of the first item. These findings lead to the conclusion that the significant decrement in C/O ratio performance due to C variation is not due to either a poorer ratio of correct responses of the first or second response item, but to the lack of correct responses between the occurrence of the first correct response and the occurrence of the second correct response. Furthermore, since the first item C/O ratio is comparable to the C/O ratio in Cond. C_1D_1 and C_1D_2, the decrement in C/O ratio must be attributed to the lack of correct responses for the second response item prior to the first correct anticipation of the second response item. This finding is quite related to the response-learning data since the slow learning of the second response item necessarily means that it cannot be associated correctly and therefore correct response frequency must be lower in the cue variation conditions.

Errors.—Columns 9 and 10 of Table 1 present the mean intralist and extralist intrusions, respectively. An analysis of variance performed upon the intralist intrusion data revealed a significant C source of variation, F (1, 56) = 4.06, $p < .05$, and a significant D source of variation, F (1,56) = 5.05, $p < .05$. The C $\times$ D source of variation is not significant. An analysis of variance performed upon the extralist intrusion data revealed a significant D source of variation,

[3] The data of one S in each group are not considered in the analysis because the same cue never occurred on the n and $n + 1$ trials.

$F(1, 56) = 12.01$, $p < .01$. The C and C $\times$ D sources of variation are not significant.

The intralist and extralist intrusion data were divided into errors prior to and after the first correct response. Analyses of variance revealed a significant D source of variation for intralist intrusions prior to the first correct response, $F(1, 56) = 9.3$, $p < .01$, extralist intrusions prior to the first correct response, $F(1, 56) = 10.4$, $p < .01$, and extralist intrusions after the first correct response, $F(1, 56) = 6.26$, $p < .01$. No other sources of variation are significant.

The intralist and extralist intrusion results therefore indicate: (*a*) pairing direction variation yielded significantly poorer performance attributable to intralist and extralist intrusions; (*b*) cue variation had an effect of borderline significance.

Conclusion.—The following conclusions with respect to cue variation appear warranted from the above data: (*a*) use of both items of a paired associate as anticipatory cues yields a decrement in correct responses; (*b*) this decrement is directly related to the difficulty of response learning of the second item of the pair; (*c*) associative learning of the first response item of the pair is retarded, but this loss is negated when opportunity for pairing is considered; (*d*) associative learning of the second response item is quite rapid. In addition, it may be concluded that pairing direction had a small, but significantly detrimental effect upon correct response performance that was related to intralist and extralist intrusions. This finding suggests that pairing direction variation produces difficulty in item discriminability.

The present results are not directly comparable to the findings of Underwood and Keppel (1963) and Schilde (1963) because of differences in experimental design. However, the present findings demonstrate the independent effects of cue and pairing variations. In particular, the cue variation results show the importance of response learning in paired-associate experiments where either item of a pair is employed as the cue (Asch & Ebenholtz, 1962). In addition, the finding of rapid, although not immediate, associative learning of the second response item is in general agreement with the bidirectional association position. On the other hand, the finding of poorer correct response and error performance with bidirectional pairing presentation suggests that unidirectional and bidirectional pairing are not equivalent, and therefore the associations are not learned bidirectionally.

REFERENCES

Asch, S. E., & Ebenholtz, S. M. The principle of associative symmetry. *Proc. Amer. Phil. Soc.*, 1962, **106**, 135–163.

Battig, W. F. Scaled difficulty of nonsense-syllable pairs consisting of syllables of equal association value. *Psychol. Rep.*, 1959, **5**, 126.

Feldman, S. M., & Underwood, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

Jantz, E. M., & Underwood, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, **56**, 174–179.

Murdock, B. B., Jr. "Backward" learning in paired associates. *J. exp. Psychol.*, 1956, **51**, 213–215.

Murdock, B. B., Jr. "Backward" associations in transfer and learning. *J. exp. Psychol.*, 1958, **55**, 111–114.

Postman, L. The temporal course of proactive inhibition for serial lists. *J. exp. Psychol.*, 1962, **63**, 361–369.

Richardson, J. Comparison of S-R learning of paired-associates. *Psychol. Rep.*, 1960, **7**, 225–228.

Schilde, M. E. Unidirectional versus bidirectional learning of paired associates. Unpublished master's thesis, University of Virginia, 1963.

Underwood, B. J., & Keppel, G. Bidirectional paired-associate learning. *Amer. J. Psychol.*, 1963, **76**, 470–474.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received February 27, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 496–502

PRIOR CONTEXT AND FRACTIONAL VERSUS MULTIPLE ESTIMATES OF THE REFLECTANCE OF GRAYS AGAINST A FIXED STANDARD[1]

E. C. POULTON, D. C. V. SIMMONDS

Applied Psychology Research Unit, Cambridge, England

RICHARD M. WARREN

University of Wisconsin, Milwaukee

AND JOHN C. WEBSTER

United States Navy Electronics Laboratory, San Diego, California

850 unsophisticated Os each made multiple estimates of darkness or fractional estimates of lightness on the same gray variables against a white standard called 1 or 100, respectively. On a normal plot the median estimates were approximately linear against reflectance. The multiple estimates were linear against the fractional estimates, whereas a reciprocal relationship is predicted by transducer theory. The slope depended upon the order of the variables; the ratio of the greatest to the smallest slope ranged between 2.1 and 1.2 to 1 ($p < .01$ or better). Conclusion: unsophisticated Os can estimate sensory magnitudes systematically, but the nature and size of the units they use are determined by present and prior experimental conditions.

It was reported in a previous paper (Poulton & Simmonds, 1963) that the very first magnitude estimates of reflectance are influenced by the range of numbers available to *O*. The present experiment investigated this effect in greater detail for variables judged against a fixed standard.

By requiring different groups of unsophisticated *O*s to judge the variables in different fixed orders, it was also possible to examine the effects of prior context more thoroughly than has been done previously. Nothing appears to be known about the effect of multiple estimates upon each other. For fractional estimates it has been found that an estimate can be changed reliably ($p < .001$) by a prior fractional estimate (data from Warren & Poulton, 1960). However it is possible that such effects are transient, and cancel out when a series of variables is used. The results reported here suggest that this is not the case.

[1] Carried out at the Applied Psychology Research Unit, Cambridge, England. The research was supported by the British Medical Research Council, by the United States Air Force Office of Scientific Research of the Air Research and Development Command through its European Office, and by the National Science Foundation. We are also grateful to the authorities and student volunteers of the Cambridge colleges in which the experiments were conducted, and to P. Freeman for statistical advice.

METHOD

Observers.—These were mainly male undergraduates of Cambridge University, although a few graduates and women students were used. Psychology students and those who had served in previous experiments were excluded. A sample of 175 of their ages ranged from 18 to 28 with a median of 21.

Display.—A piece of white paper 17.5 × 11.3 in. with a reflectance of 83% served as a background. It lay on a table directly in front of *O*, arranged symmetrically with a long side closest to him. For the groups of *O*s making four fractional judgments of lightness it was labeled 100 in pencil in the left-hand corner; for all multiple judgments of

darkness it was labeled 1. In the middle of the white paper was placed a gray stimulus patch 2.0 × 1.5 in. with its long side parallel to the long side of the background. The percentage reflectance of the patch was 68, 58, 44, 19, 7.8, or 3.4. The white paper background and small patches were coated with nonglossy paint. Their reflectances were measured both with a Macbeth illuminometer and with an illuminometer made by Salford Electrical Instruments Limited. As previously (Poulton & Simmonds, 1963), the lighting on the table was at the ambient level to which *O* had become adapted.

Experimental conditions.—Each *O* judged in turn two, three, or four grays. The reflectances and the orders of presentation of the groups making four judgments are shown in Table 1. The same reflectances and orders were used for both the multiple and fractional estimates; thus the multiple estimates of Group A correspond to the fractional estimates of Group E, and so on. Each *O* made either multiple or fractional estimates, never both.

Procedure.—The experiment was conducted in Cambridge Colleges. One *E* canvassed for volunteers and kept curious eyes off the display. When *O* sat down at the table, the second *E* handed him a white card 8 × 5 in. with instructions typed on it, a pencil, and a white answer slip 4 × 3 in. For fractional judgments of lightness the instructions read: "You will see a gray patch of paper lying on a white blotter. If the lightness of the white blotter is called 100, what (smaller) number describes the lightness of the gray patch? Please write down your answer on the slip of paper provided." The instructions were explained if necessary by paraphrasing them. For multiple judgments of darkness "white" was changed to "light gray," "lightness" to "darkness," "100" to "1," and "smaller" to "larger." (The change to "light gray" was made to satisfy those *O*s who were unwilling to accept a number other than zero for "white.")

As soon as *O* had written down his first numerical estimate, the second *E* changed the stimulus patch and asked him to write his second answer below his first. The third and fourth answers were written on the reverse side of the answer slip. Some of the *O*s who made only two judgments were asked their age. Finally *O* was requested not to discuss his answers with anyone who had not already done the experiment. This procedure was repeated with successive volunteers until a group of 50 acceptable answers had been obtained. Answers which did not satisfy the instructions were discarded.

Calculations.—Two distinct median measures were calculated. The *group median* plotted in the figures was the median of all the *O*s in the group. They generally numbered 50, but in Fig. 2b some of the medians were computed from the pooled scores of two, or in one case three, groups of 50 *O*s. The *mean median* of Table 1 was calculated from the medians of the five subgroups comprising the first and last five *O*s in the group, the second and one-from-last five *O*s, and so on. Its 5% fiducial limits are an index of the lack of precision in the median which is tolerated when using experimental groups of only 10 *O*s. The straight lines in the figures were fitted by the method of least squares. This method was also followed in calculating the quadratic terms. Two-tailed tests have always been used.

Results

The group medians, mean medians, and their 5% fiducial limits for the groups of 50 *O*s which made four multiple estimates of darkness are shown in the top half of Table 1. Figure 1a shows the group medians plotted against reflectance on a normal plot, while Fig. 1b is the corresponding log-log plot. The straight lines in Fig. 1a were fitted by the method of least squares. Only the quadratic term of Group B was reliable, and here reliability reached only the .05 level of significance. No attempt was made to fit the points in Fig. 1b by straight lines. The functions for the reciprocals of the medians (not shown) were also markedly curvilinear on a log-log plot.

The slopes of the lines in Fig. 1a, together with their 5% fiducial limits, are given in the one from last column of Table 1. The slope of each group appears to have been determined by the reflectance of the very first variable. The slope of Group A which judged first the variable of 68% reflectance, was reliably smaller than the slopes of Groups B and C which

TABLE 1

ORDER OF VARIABLES

Group	First Variable			Second Variable			Third Variable			Fourth Variable			Fitted Slope	
	% Reflectance	Numerical Estimate		% Reflectance	Numerical Estimate		% Reflectance	Numerical Estimate		% Reflectance	Numerical Estimate			
		Group Mdn.	Mean Mdn.		Group Mdn.	Mean Mdn.		Group Mdn.	Mean Mdn.		Group Mdn.	Mean Mdn.	Normal Plot	Log-Log Plot
Multiple Estimates of Darkness														
A	68	4	4.2±.8	44	8	7.9±2.2	7.8	14	14.9±7.0	58	5	5.3±1.4	−.17±.01[a]	
B	44	9	8.9±1.2	68	4	3.7±.9	7.8	19.9	16.4±5.4	58	5	5.8±1.5	−.23±.03[a]	
C	44	9.7	12.4±12.4	7.8	20	25.4±17.9	68	4	3.8±2.3	19	14.5	16.1±10.0	−.23±.02[b]	
D	7.8	25	36.6±29.9	44	10	11.3±9.5	68	4.8	6.2±5.3	19	24.5	27.8±22.0	−.32±.04[b]	
Fractional Estimates of Lightness														
E	68	74.8	73.5±4.4	44	50	50.1±3.3	7.8	10.1	12.1±5.1	58	70	71.2±2.7	1.23±.08[c]	.98±.05[c]
F	44	60	60.5±5.6	68	80	81.3±6.3	7.8	10	9.6±4.0	58	70	71.2±8.3	1.17±.07[d]	.96±.04[d]
G	44	59.8	57.6±8.3	7.8	24.4	20.9±8.7	68	85	85.7±5.0	19	40	43.2±4.4	.98±.04[d]	.61±.04[d]
H	7.8	20	19.1±5.7	44	65	65.4±9.5	68	85	84.8±5.7	19	27	28.9±8.2	1.06±.06[c]	.73±.08[c]

Note.—For the multiple estimates of darkness the standard of 83% reflectance was called 1, while for the fractional estimates of lightness it was called 100.

[a] Group A–Group B, $p < .05$.
[b] Group C–Group D, $p < .05$.
[c] Group E–Group H, $p < .05$.
[d] Group F–Group G, $p < .05$ or better.

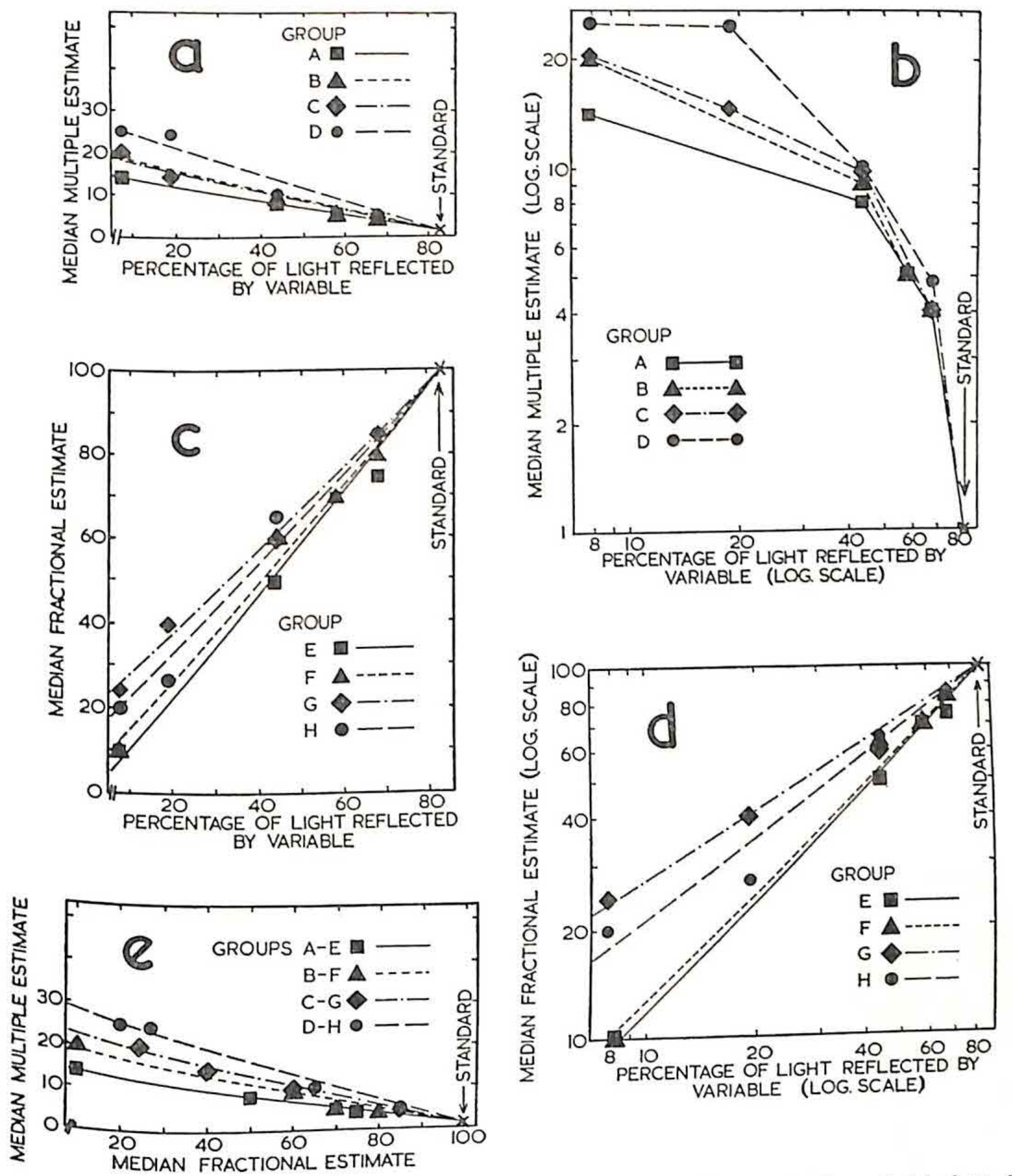


FIG. 1. Median numerical estimates of groups of 50 *O*s. (Figures 1a through 1d plotted against the reflectance of the variable; Fig. 1e multiple estimates plotted against the fractional estimates of different *O*s judging the same variables in the same order. Except for Fig. 1b, a straight line passing through the standard has been fitted to the data of each group or pair of groups by the method of least squares.)

judged first the variable of 44% reflectance. And these in turn were reliably smaller than the slope of Group D which judged the variable of 7.8% reflectance first. The ratio of the slopes of Groups D and A was over 1.8 to 1.

The corresponding data for fractional estimates are given in the bottom half of Table 1 and in Fig. 1c and d. In both the normal and the log-log coordinates the fitted straight lines shown left quadratic terms which were unreliably statistically. Here

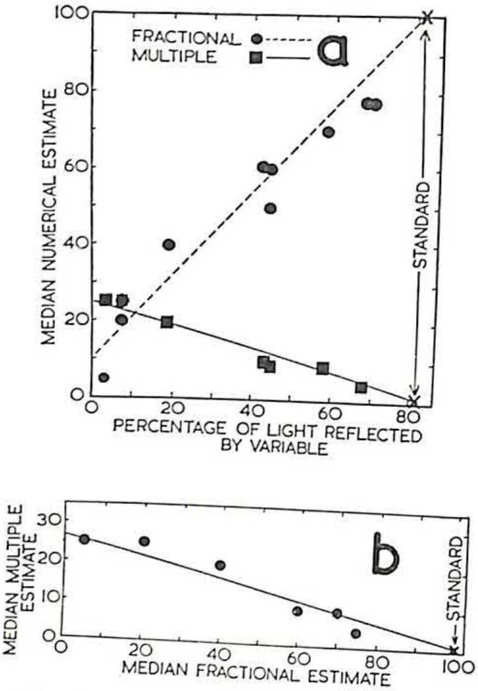


FIG. 2. Median numerical estimates of the very first judgments of groups of at least 50 *Os* plotted against the reflectance of the variable (Fig. 2a); and multiple estimates plotted against fractional estimates (Fig. 2b). (The straight lines passing through the standard have been fitted by the method of least squares.)

the slopes, which are given in the last two columns of Table 1, appear to depend upon the reflectance of the first variable which was not near the middle of the range of reflectances. In both sets of coordinates the slope of Group E which was presented first with the variable of 68% reflectance, was reliably greater than the slope of Group H which judged the variable of 7.8% reflectance first. Of the two groups which first judged the variable of 44% reflectance, Group F which followed this by the variable of 68% reflectance had a reliably greater slope than Group G which judged second the variable of 7.8% reflectance. The ratio of the slopes of Groups E and G was over 1.2 to 1 on the normal plot, and over 1.6 to 1 on the log-log plot. In both cases it was in the opposite direction to the ratio of the slopes of Groups A and D.

Figure 1e shows the median multiple estimates plotted against the median fractional estimates of the groups which judged the same variables in the same order. For each pair of groups the straight line fitted to the data left a quadratic term which was not reliable ($p > .05$). The slope of the combined Groups D and H, $-.30 \pm .02$, was reliably the largest, and the slope of Groups A and E, $-.14 \pm .01$, was reliably the smallest ($p < .02$ or better in each case). The ratio of these two slopes was over 2.1 to 1. When the reciprocals of the median multiple estimates were plotted against the fractional estimates and vice versa, the functions (not shown) were all markedly curvilinear with both normal and log-log coordinates.

Figure 2 shows the corresponding data for the very first judgments of groups of at least 50 *Os*. The straight line in Fig. 2a fitted to the median multiple estimates had a slope of $-.29 \pm .01$. The quadratic term was just reliable ($p < .05$), indicating a slight concavity upwards as for Group B. On a log-log plot the function was markedly curvilinear, resembling those in Fig. 1b.

The straight line fitted to the median fractional estimates had a slope of $1.08 \pm .03$ on the normal plot of Fig. 2a, and a slope of $.76 \pm .05$ on a log-log plot (not shown). In both cases the quadratic term was not reliable ($p > .05$). Figure 2b shows the linear relationship between the very first multiple and the very first fractional estimates. The fitted line had a slope of $-.27 \pm .02$; the quadratic term was not reliable.

Discussion

Linear relationships on a normal plot.—Figures 1a, 1c, and 2a show that the estimates of both darkness and lightness were approximately linear when plotted against reflectance on a *normal* plot. The results of fractional estimates of lightness are distantly related to the data of Stevens and Galanter (1957, Fig. 12) and of Torgerson (1960). For the slopes are all fairly close to unity; thus points lying close to one of the straight lines also lie close to a straight line on the corresponding log-log plot of Fig. 1d. The fitted exponents of Fig. 1d, which range from .61 ± .04 to .98 ± .05, are probably reliably smaller than the value reported by Stevens and Galanter and by Torgerson of 1.2, although these authors give no confidence limits. But the higher value can perhaps be attributed to the multiple estimates which their *O*s made as well as fractional, since prior multiple estimates would be expected to increase the size of the exponent for fractional estimates (Poulton & Simmonds, 1963).

In contrast the results for multiple estimates of darkness cannot be related to previous results, since they are by no means linear when plotted against reflectance on a log-log plot (Fig. 1b). Nor do the reciprocals of the medians, which on transducer theory (Stevens, 1961a, p. 27) should correspond to fractional estimates of lightness, lie on a straight line when plotted against reflectance on either a normal or a log-log plot.

Effect of prior context.—The differences in slope in Fig. 1a and c suggest that prior context can have considerably more general and more pronounced effects on magnitude estimates than has so far been realized (Stevens, 1956; Stevens & Poulton, 1956). The effects shown were not due simply to selecting a first variable too far to one end or other of the range of variables. For Groups F and G both started with a variable of 44% reflectance, which is near the middle of the range; yet their slopes were reliably different on both the normal and the log-log plot (Table 1). Nor were the effects due simply to the particular selection of variables, for the reflectances covered most of the available range. Also Groups A and B actually had an identical set of variables to judge, as did Groups C and D, yet in each case the fitted slopes were reliably different. Figures 1a and c show that in general the fits of the straight lines were good. Thus the groups were reasonably consistent in the differences which they had acquired.

Table 1 shows that for the multiple estimates of darkness the ratio of the largest slope to the smallest was over 1.8 to 1. For the fractional estimates of lightness the ratio was over 1.2 to 1 on the linear plot, and over 1.6 to 1 on the log-log plot. Differences in slope of this order of size cast more doubt upon the representativeness of the slopes for different sensory continua obtained by magnitude estimation (Stevens 1960a, 1960b, 1961a, 1961b) especially when the order in which the variables were judged is not stated.

Effect of number scale.—On a simple transducer theory (Stevens, 1961a, p. 27) the multiple estimates of darkness should have borne a reciprocal relationship to the fractional estimates of lightness, since the stimuli were identical. Figures 1e and 2b show that instead the relationship was linear. Apparently if *E* presents unsophisticated *O*s with a range of reflectances and a number scale, they will arrange the reflectances more or less linearly along the number scale whatever form the number scale takes. This allocation of numbers to reflectances is systematic (see Fig. 1a, 1c, and 2a), but it cannot be said to be "magnitude estimation" when as here one number scale is related reciprocally to the other. Clearly *O*s must be able to appreciate sensory magnitudes in order to be systematic. But it would appear that a "magnitude estimation" experiment does not necessarily indicate the nature of *O*s' units of judgment, since they are a function of the number scale *E* tells them to use.

That this has not been reported previously must be because no *E* has ever required unsophisticated *O*s to make a

number of multiple estimates without also making or having made fractional estimates. And a multiple estimate following a fractional estimate can be reliably different from a multiple estimate made as the very first judgment of magnitude (Poulton & Simmonds, 1963).

REFERENCES

Poulton, E. C., & Simmonds, D. C. V. Value of standard and very first variable in judgments of reflectance of grays with various ranges of available numbers. *J. exp. Psychol.*, 1963, **65**, 297–304.

Stevens, S. S. The direct estimation of sensory magnitudes-loudness. *Amer. J. Psychol.*, 1956, 69, 1–25.

Stevens, S. S. On the new psychophysics. *Scand. J. Psychol.*, 1960, **1**, 27–35. (a)

Stevens, S. S. Ratio scales, partition scales and confusion scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960. Ch. 6. (b)

Stevens, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Ch. 1. (a)

Stevens, S. S. Toward a resolution of the Fechner-Thurstone legacy. *Psychometrika*, 1961, **26**, 35–47. (b)

Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

Stevens, S. S., & Poulton, E. C. The estimation of loudness by unpracticed observers. *J. exp. Psychol.*, 1956, **51**, 71–78.

Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960. Ch. 3.

Warren, R. M., & Poulton, E. C. Basis for lightness-judgments of grays. *Amer. J. Psychol.*, 1960, **73**, 380–387.

(Received March 2, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 503–510

LOUDNESS, A PRODUCT OF VOLUME TIMES DENSITY[1]

S. S. STEVENS, MIGUELINA GUIRAO, AND A. WAYNE SLAWSON

Laboratory of Psychophysics, Harvard University

Experiments were designed to determine the relation between loudness and 2 other auditory attributes, volume (apparent size) and density (apparent compactness or concentration). 2 sets of stimuli, quarter-octave bands of noise covering a wide range of center frequencies and SPL, were presented through earphones to Os who made magnitude estimations of one or another of the attributes. The loudness estimations were plotted against loudness level and found to agree with the sone scale. A plot of the estimations of loudness against the product of the estimations of volume times the estimations of density produced a slope of 1.0 in log-log coordinates. Loudness is therefore proportional to volume times density. This relation was confirmed by experiments involving magnitude estimations of the inverse attributes, softness, smallness, and diffuseness. These inverse attributes were found to be the reciprocals of their respective direct attributes. As predicted, softness turned out to be proportional to the product of smallness times diffuseness. Thus magnitude estimations of both the attributes and their inverses established the proportionality between loudness and the product of volume times density.

One of the earliest criticisms to greet the experimental demonstration (Stevens, 1934) that *O*s can judge the density of tones was the conjecture that tonal volume (apparent size or largeness) and tonal density (compactness or concentration) are merely the inverse of each other and not two different attributes. It was noted that when the frequency is raised volume decreases and density increases, which suggests a reciprocal or inverse relation. But both volume and density increase with stimulus intensity, which seems to deny the inverse relation. Clearly the problem is more complex than was implied by some of the early discussion.

The successful scaling of the two auditory attributes, volume (Terrace & Stevens, 1962) and density (Guirao & Stevens, 1964), reopened the question of the interrelations among the attributes. Those studies showed that both volume and density grow as power functions of sound pressure. The point of interest in the present context is the behavior of the exponents of the power functions. The exponents of the volume functions grew larger with frequency and the exponents of the density functions grew smaller. There was, in fact, a suggestion that at each frequency the two exponents may add approximately to a constant sum, a sum not greatly different from 0.6, the exponent of the loudness function.

A decision was therefore made to explore more directly the relations among volume, density, and loudness. What appeared to be needed was a series of experiments in which the stimuli would be narrow bands of noise and the same set of stimuli would be used throughout a given series of experiments. Bands of noise were chosen over pure tones mainly because *O*s seem able to judge noises with greater certainty than tones. Pure

[1] This research was supported in part by the National Science Foundation (Grant G-10716) and in part by the National Institutes of Health (Grant B-2974). This is Laboratory of Psychophysics Report PPR-292.

TABLE 1

Geometric Means (*GM*) and Quartile Deviations (*QD* in Decilogs) of the Magnitude Estimations of Three Attributes of Narrow Bands of Noise

Noise Bands		Loudness		Volume		Density	
Freq.	SPL	*GM*	*QD*	*GM*	*QD*	*GM*	*QD*
Series A							
200	55	1.5	1.5	26.6	3.5	1.3	1.5
325	86	31.3	0.8	51.2	0.7	11.4	2.3
450	68	14.1	1.5	19.9	1.5	5.4	1.5
653	51	3.6	1.1	12.3	1.9	2.5	1.5
653	100	76.5	1.4	59.8	2.0	33.7	1.0
1,000	75	20.9	1.5	13.2	2.0	17.7	0.6
1,630	89	39.9	1.8	24.4	3.0	36.3	2.7
2,000	45	3.5	1.8	4.7	1.1	6.4	2.0
4,000	95	77.7	1.8	15.2	4.0	87.6	2.0
6,000	63	5.8	2.0	3.8	1.1	28.4	2.4
Series B							
200	50	1.0	1.5	18.3	2.0	0.8	2.2
325	80	23.3	2.0	27.7	1.6	9.2	2.2
720	100	79.4	1.1	38.0	2.7	51.0	1.1
720	42	1.9	0.9	8.1	2.0	2.1	0.3
1,000	70	12.9	1.1	11.7	0.3	13.1	0.4
2,580	90	64.2	1.5	12.5	3.3	60.4	0.8
3,000	50	6.3	1.4	4.6	0.9	8.8	1.8
6,000	70	12.6	1.1	2.9	2.0	21.7	1.9
Series C							
200	50			12.0	2.9	1.5	1.5
325	80			25.2	2.4	8.6	2.4
720	100			45.0	1.5	40.0	1.2
720	42			4.5	2.7	3.4	1.5
1,000	70			9.9	0.5	15.6	1.1
2,580	90			18.2	1.5	61.6	1.1
3,000	50			3.2	1.1	16.1	2.0
6,000	70			3.2	3.0	40.8	0.9

Note.—The center frequency and sound pressure level are given for each band. A standard at 910 cps and 65 db. SPL was presented at the beginning of each run and called "10." No loudness judgments were made in Series C.

tones have an indefiniteness about them that makes even their loudness less easy to judge than the loudness of a band of noise (Stevens, 1956).

In addition to the scaling of loudness, volume, and density by magnitude estimation, the inverse of each of the attributes was scaled. If the same functional relation among the attributes can be established by measuring the reciprocals as well as the attributes themselves, the nature of the relation becomes more securely established.

Method

The procedure was essentially the same as that followed in the earlier experiments on density (Guirao & Stevens, 1964). Bands of noise were generated by filtering a white noise with the aid of a continuously variable, constant K, filter network (Allison 2B). For a given stimulus the high- and low-pass sections of the filter were set to the same frequency, which produced a band about a quarter of an

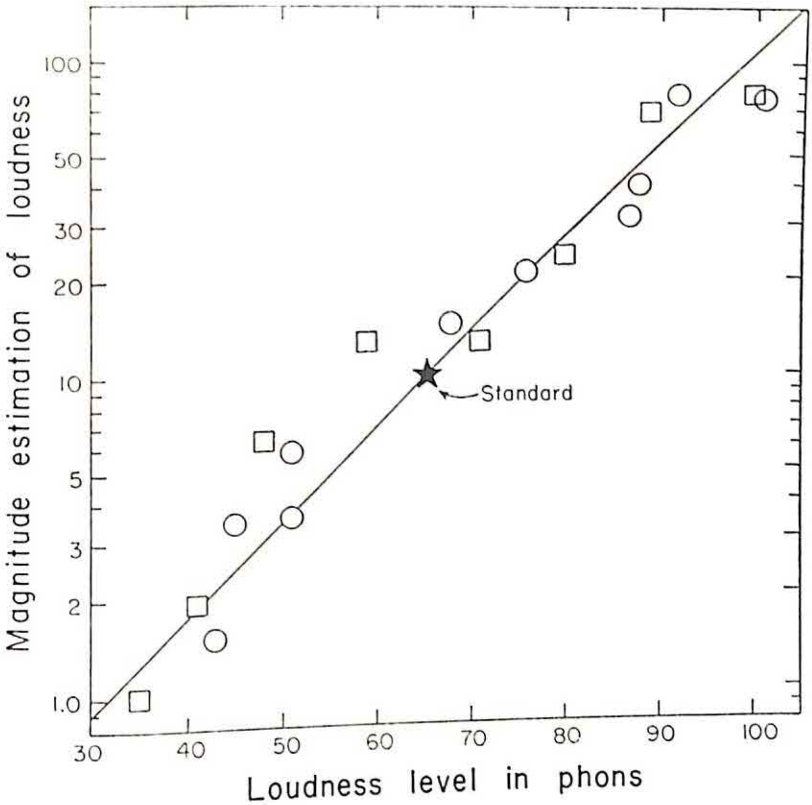


FIG. 1. Relation of loudness estimations to loudness level. (The squares and circles stand for two different experiments. The ordinate of each point is the geometric mean of 20 estimations, 2 by each of 10 *O*s. The abscissa of each point was determined by converting the stimulus frequency and SPL [Table 1] into loudness level with the aid of equal-loudness contours. The line through the data shows the slope [exponent] predicted by the standard sone scale.)

octave wide at a level 3 db. down from the peak. The stimuli were delivered through PDR-8 earphones mounted in sponge neoprene cushions MX-41/AR. The calibration of the two earphones in a 6-cc coupler showed them to have a reasonably flat response characteristic over the range of frequencies of interest. The voltage across the two earphones connected in series was read with a Ballantine root-mean-square VT voltmeter, Model 320, and converted into sound pressure level (SPL) on the basis that 1 v. corresponds to 100 db.

Ten *O*s chosen more or less at random from among 30 students and staff of the laboratory served in each experiment. The stimuli were presented twice each in irregular order, a different order for each *O*. A band of noise centered at 910 cps and 65 db. SPL was presented at the outset of each experiment and assigned the value 10. The *O* was allowed to hear the standard again whenever he requested it. Two sets of stimuli varying widely in both frequency and SPL were used. The center frequencies and the SPL of the bands are tabulated in Table 1.

The instructions were as follows, with appropriate variations for the different attributes—loudness, volume, or density—and their inverses—softness, smallness, and diffuseness.

> I am going to present a series of noises. Your task is to judge the loudness of each noise. The loudness of the first noise will be called 10. Assign to each of the succeeding noises a number proportional to its apparent loudness, remembering that the loudness of the first noise was called 10. For example, if the second noise sounds four times as loud, call it 40; if half as loud, call it 5, and so forth.

Further definitions were also included in the instructions for most of the attributes. Volume was said to refer to the size of the sound, i.e., how large or small it appears to be. Density was said to refer to the compactness,

concentration, or hardness of the sound. Diffuseness was said to refer to how tenuous, thin, or rarefied the sound appears to be

Results

Loudness

A first concern is whether *O*s can make satisfactory judgments of loudness when the stimuli vary in frequency as well as in sound pressure. Two separate experiments were run, each with a different set of stimuli. The geometric means of the magnitude estimations are shown in Table 1, Series A and B. In order to compare these data with the standard loudness function, each SPL was converted to its corresponding loudness level by means of the equal-loudness contours presented by Stevens and Davis (1938). These contours are based on the equal-loudness data of Fletcher and Munson combined with the curve for minimum audible pressure determined by Sivian and White. Although the contours are for pure tones, they are presumably applicable to narrow bands of noise whose width is of the order of a critical band (Zwicker, Flottorp, & Stevens, 1957).

Figure 1 shows the loudness estimates plotted against loudness level in phons. The line through the data has the slope (exponent) of the standard "sone" scale. Although a few of the points deviate fairly far from the line, the overall picture provides an interesting confirmation of the loudness function, for *O*s were required to judge loudness in the face of wide variations in other aspects of the stimuli. Part of the variability must, of course, be attributed to imperfections in the equal-loudness contours that were used to determine loudness level. Some of the variability must also be due to the inevitable uncertainties regarding the true sound pressure levels in the ears of the listeners (Beranek, 1949, p. 738).

The data agree with the sone scale of loudness somewhat more closely if the equal-loudness contours presented by Stevens and Davis are altered so as to preserve their upward concavity throughout the entire intensity range. The agreement then is quite good, and it suggests that the equal-loudness contours could be mapped out directly by magnitude estimation if a sufficiently thorough experiment were run.

Density and Volume

Three experiments were run to scale each of the attributes, volume and density. The geometric means of the magnitude estimations are listed in Table 1. In order to verify that the data for density were similar to those obtained in the earlier study (Guirao & Stevens, 1964) with bands of noise, the magnitude estimations were plotted against SPL with frequency as a parameter. The technique employed was that used in the earlier study and the outcome was essentially the same. Density was found to grow as a power function of sound pressure. The exponent decreases as frequency increases. Only at the frequency 250 cps was there a suggestion of a slightly different exponent from that found in the earlier study. Consequently, this test of the adequacy of the data appeared to be reasonably well satisfied.

Precisely the same test could not be applied to the volume judgments because the volume functions have not been determined for bands of noise. It was nevertheless possible to compare the data for volume in Table 1 with the volume functions for pure tones presented by Terrace and Stevens (1962). There is generally good agreement, but with occasional poor agreement at the highest frequency tested. It was typically the frequencies in the region above 3,000 cps that caused the most difficulty in all aspects of these experiments.

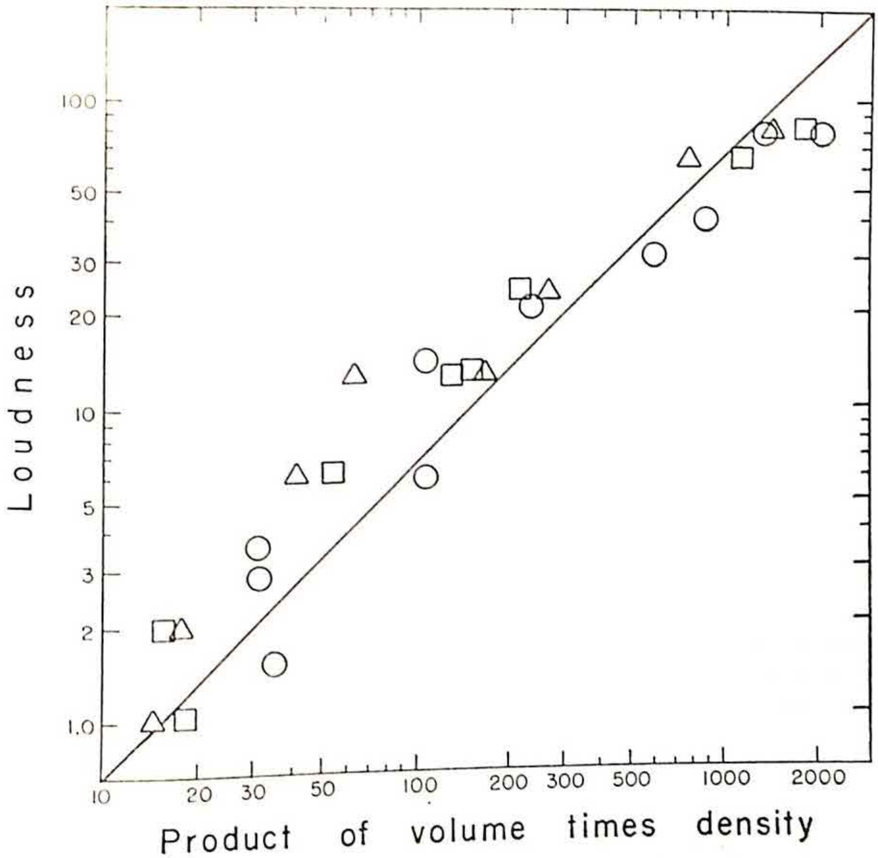


FIG. 2. Approximate proportionality between loudness and the product of volume times density. (Each point represents a different product of volume times density, but the squares and triangles represent the same set of loudness values. Exact proportionality is indicated by the line whose slope in the log-log coordinates is 1.0.)

Loudness vs. Volume Times Density

For each stimulus the geometric mean of the magnitude estimation of volume was multiplied by the corresponding estimation of density and plotted on the abscissa in Fig. 2. On the ordinate in Fig. 2 are plotted the magnitude estimations of loudness. Thus both coordinates are psychophysical attributes.

If auditory density is proportional to loudness divided by volume, as the analogy with physical density might suggest, the data in Fig. 2 should describe a slope of 1.0 in the log-log coordinates. That is the slope of the line actually drawn. The fit appears sufficiently good to warrant the conclusion that loudness is proportional to volume times density. The most deviant point corresponds to one of the high-frequency stimuli, 6,000 cps.

Inverse Functions

It has been demonstrated on at least seven perceptual dimensions that *O*s can make magnitude judgments of the inverse of a continuum (Stevens & Guirao, 1963; Torgerson, 1960). The *O*s produce thereby a function that approximates the reciprocal of the power function obtained when the continuum itself is judged. This procedure of inverse judgments promised a means by which the validity of the relation shown in Fig. 2 could be checked by an independent approach. The strategy was to scale each of the

TABLE 2

GEOMETRIC MEANS (*GM*) AND QUARTILE DEVIATIONS (*QD* IN DECILOGS) OF THE MAGNITUDE ESTIMATES OF THREE INVERSE ATTRIBUTES OF NARROW BANDS OF NOISE

Noise Bands		Softness		Smallness		Diffuseness	
Freq.	SPL	*GM*	*QD*	*GM*	*QD*	*GM*	*QD*
720	100	1.1	1.5	2.0	1.5	4.6	2.9
2,580	90	1.2	2.0	4.0	2.0	2.1	3.0
325	80	5.3	1.2	4.2	1.6	16.5	1.5
1,000	70	7.0	0.8	9.9	0.5	9.8	0.9
6,000	70	8.9	0.9	31.0	2.0	2.8	2.0
3,000	50	16.7	1.3	28.5	0.8	8.4	2.0
720	42	42.0	1.3	10.9	3.0	36.0	1.9
200	50	63.0	1.7	4.3	3.4	59.0	1.7

Note.—The center frequency and sound pressure level are given for each band. A standard at 910 cps and 65 db. SPL was presented at the beginning of each run and called "10."

inverse continua, softness, smallness, and diffuseness, and to test the proportionality between softness and the product of smallness and diffuseness.

The stimuli and the geometric means of the magnitude estimations are listed in Table 2.

In order to examine the degree to which the inverse judgments approximate the reciprocal of the direct judgments, the two sets of geometric means for each continuum were plotted in Fig. 3. The lines with slopes of -1 indicate the predicted slope of the inverse functions on the assumption that they obey the reciprocal law.

The magnitude estimations of softness are almost the reciprocal of the estimations of loudness. There is a slight tendency for the function, softness vs. loudness, in Fig. 3 to be concave downward, a feature that has characterized many experiments on inverse scaling. The reason for this curvature is not understood.

Evidence of a similar slight curvature is apparent in the plot of diffuseness vs. density, although the reciprocal relation is closely approximated.

Volume (largeness and smallness) fared less well on the whole than did the other attributes. The estimates of smallness fall fairly close to the reciprocal line, but there is a general constriction of the overall range of the judgments. The fact of the matter is that *O*s reported more trouble with volume, both largeness and smallness, than they reported with loudness or density. This fact does not necessarily mean that volume is the most difficult attribute to judge. It may be relevant that most of the *O*s had served in experiments on density during the months prior to the present series of experiments. It is conceivable that a practiced familiarity in the abstracting of the density attribute made it difficult to switch to volume.

In any case, the goodness of the fit to the reciprocal functions in Fig. 3 suggests that the inverse judgments should be adequate to test the relation between softness and the product of smallness times diffuseness. Accordingly, the geometric means of the inverse judgments were plotted in Fig. 4. The line in Fig. 4 has the slope 1.0 predicted by the hypothesis. The point that departs farthest from the line represents the stimulus whose frequency, 200 cps, was the lowest used and the one that produced the largest variability in the judgments of smallness. The other points in Fig. 4 are satisfactorily close to the predicted line.

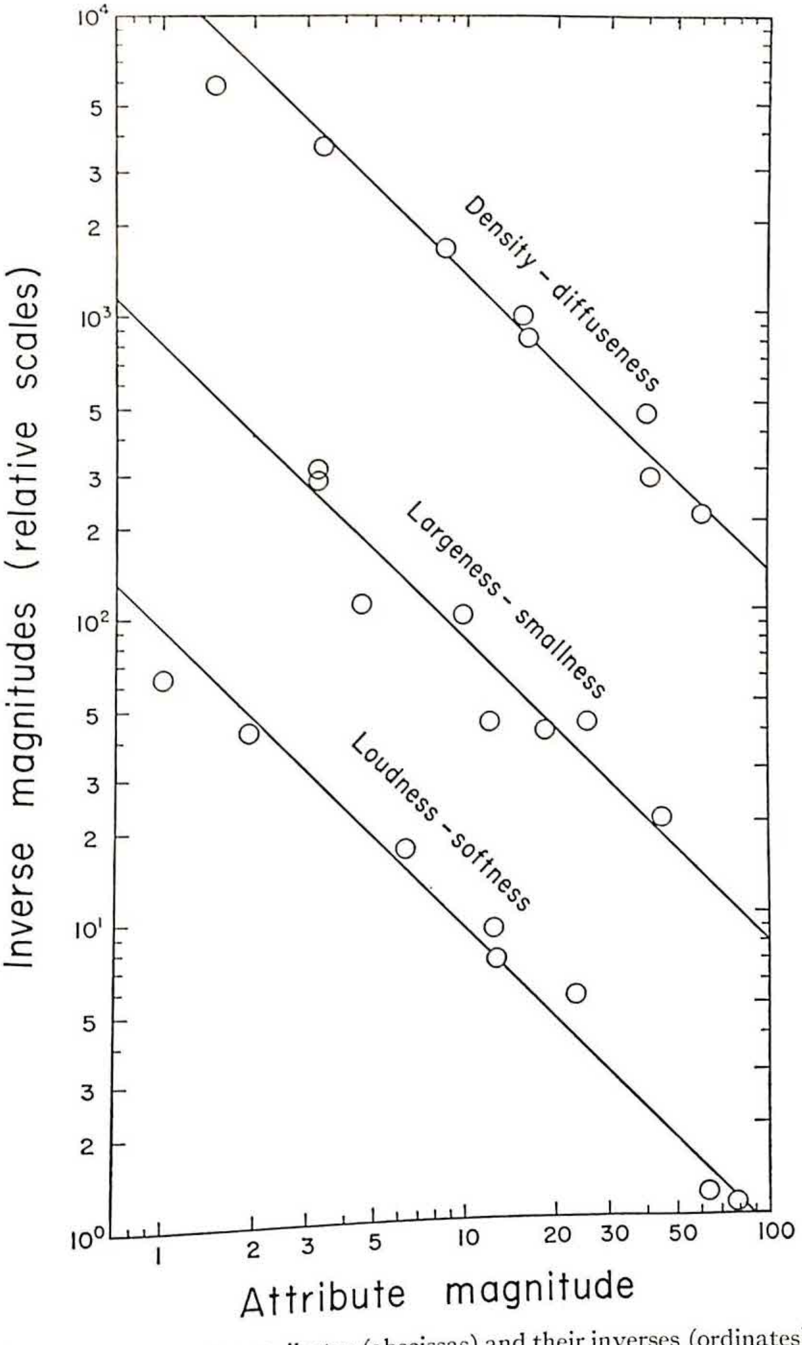


FIG. 3. Relation between the attributes (abscissas) and their inverses (ordinates). (Exact reciprocality is indicated by the lines having slopes of -1 in the log-log plot. Data are from Tables 1 and 2.)

Variability

The quartile deviations (*QD*) were measured by taking half of the range that included the centermost 10 among the 20 magnitude estimations for each stimulus. The values entered in Tables 1 and 2 are expressed in decilogs, a tenth of a common logarithmic unit. The justification for logarithmic measures lies in the symmetry of the logarithmic distributions. It is this same symmetry that calls

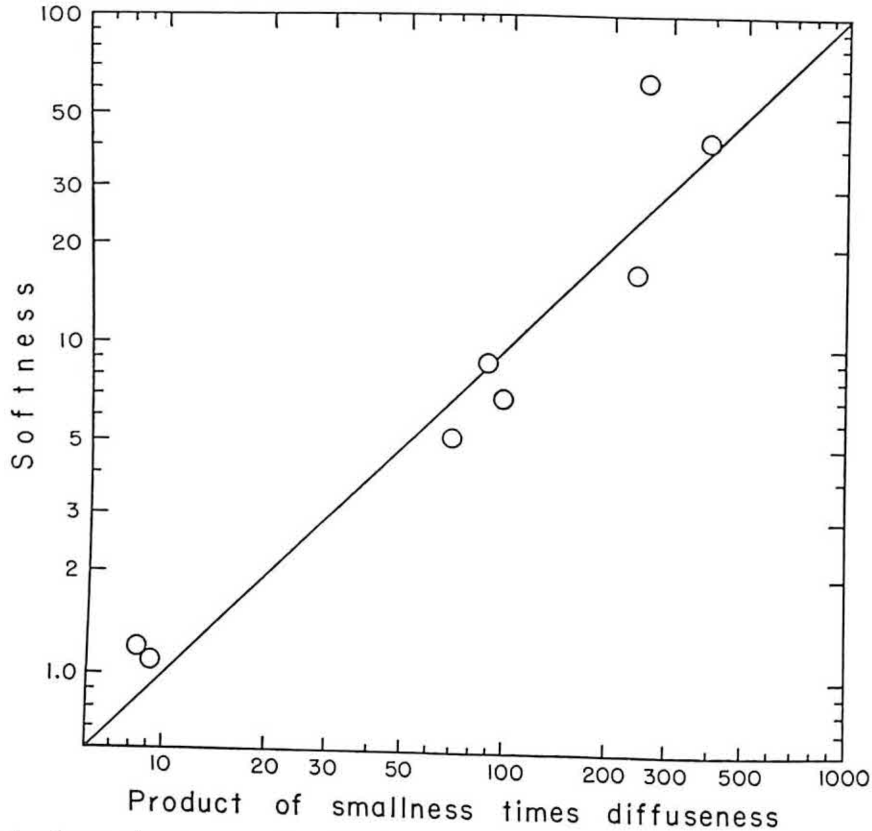


FIG. 4. Approximate proportionality between softness and the product of smallness times diffuseness. (This relation is predicted by the function in Fig. 2. Exact proportionality is indicated by the straight line with a slope of 1.0.)

for the geometric rather than the arithmetic mean in the averaging of magnitude estimations.

The medians of all the quartile deviations show that the variability tended to be smaller for judgments of loudness (1.5 decilogs) and density (1.5 decilogs) than for judgments of volume (2.0 decilogs).

The median of the quartile deviations for the judgments of softness (1.3 decilogs) was smaller than that for the judgments of smallness (1.8 decilogs) and diffuseness (1.9 decilogs).

REFERENCES

BERANEK, L. L. *Acoustic measurements.* New York: Wiley, 1949.

GUIRAO, M., & STEVENS, S. S. The measurement of auditory density. *J. Acoust. Soc. Amer.*, 1964, **36**, 1176–1182.

STEVENS, S. S. Tonal density. *J. exp. Psychol.*, 1934, **17**, 585–592.

STEVENS, S. S. Calculation of the loudness of complex noise. *J. Acoust. Soc. Amer.*, 1956, **28**, 807–832.

STEVENS, S. S., & DAVIS, H. *Hearing: Its psychology and physiology.* New York: Wiley, 1938.

STEVENS, S. S., & GUIRAO, M. Subjective scaling of length and area and the matching of length to loudness and brightness. *J. exp. Psychol.*, 1963, **66**, 177–186.

TERRACE, H. S., & STEVENS, S. S. The quantification of tonal volume. *Amer. J. Psychol.*, 1962, **75**, 596–604.

TORGERSON, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling.* New York: Wiley, 1960. Pp. 21–31.

ZWICKER, E., FLOTTORP, G., & STEVENS, S. S. Critical band width in loudness summation. *J. Acoust. Soc. Amer.*, 1957, **29**, 548–557.

(Received March 2, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 511–514

EFFECT OF STIMULUS VARIABLES ON CHOICE REACTION TIMES AND THRESHOLDS [1]

J. KASWAN AND S. YOUNG

University of California, Los Angeles

For 8 different groups of Ss, individual patterns from sets of either 4-, 3-, or 2-alternative patterns were shown to each S under 2 conditions in a forced-choice recognition task. In one condition, exposure duration was varied by E and thresholds were determined for each pattern. In the other condition, choice reaction times (CRT) were obtained, measured from stimulus onset to S's key press. It was predicted and found that the order of CRTs would parallel the order of thresholds in each combination of patterns. This finding supports the assumption that the time required to receive spatial information sufficient to distinguish patterns is an identifiable component of CRT.

A major purpose of this study is to examine the role of the exposure-duration threshold in the total time required for the identification of a visual pattern, as indicated by a choice reaction time (CRT).

Previous work has shown that the probability of correctly identifying a pattern consisting of a single horizontal array of variably spaced elements is a function of increasing exposure duration. The exposure duration required for a given level of correct response to a particular pattern depends on (*a*) the size and configuration of differences between spaces within a pattern and (*b*) the spacing sequences which characterize the alternative patterns (see Kaswan, Young, & Nakamura, 1965).

In the present study, it is assumed that exposure-duration thresholds reflect the time required to obtain sufficient spacing information about a pattern to distinguish it from other patterns. The assumption that stimulus patterns must be exposed for some small but specific amount of time before they can be accurately discriminated suggests that this time should be reflected in measures other than exposure-duration thresholds. We expect that in a task in which a stimulus is exposed throughout the choice reaction time, measured from stimulus onset to *S*'s response, the function of the initial part of the exposure is largely to convey spacing information about the pattern. The initial component of the CRT thus presumably includes a time correlated with the threshold. If the differences in CRT for patterns within a set of alternatives are due largely to differences in exposure duration required to identify these patterns, then differences in CRT should parallel differences in threshold.

[1] This study was supported by a grant from the National Science Foundation (G-9589) and a Public Health fellowship for S. Young (5-F1-MH-19,664-02). We wish to thank R. Shomer for helping to collect the data, and thank Gale Kratz for helping to analyze it.

METHOD

Materials.—Four stimuli were used in the study. Each stimulus consisted of a single linear horizontal array of $\frac{1}{4}$-in. vertical lines printed on blank off-white cardboard. The Ev pattern array consisted of 26 lines spaced evenly with 6 mm. between lines. The U stimulus contained 24 lines spaced with randomly located 6-, 8-, 7-, and 5-mm. spaces. The 3 stimulus contained 24 lines spaced in the sequence 5, 5, 7, 5, 5 mm. to appear as groups of three lines. The 1,2 stimulus con-

sisted of 24 lines spaced in the sequence 6, 4, 8, 6, 4 mm. (see Kaswan et al., 1965, for samples of the Ev, 3, and 1,2 stimuli).

Subjects.—A total of 146 men and women college freshmen and sophomores was randomly divided into eight groups. A four-alternative group of 30 *S*s was presented with the four stimuli described above. A three-alternative group of 20 *S*s received the Ev, 3, and 1,2 patterns. There were six two-alternative groups of 16 or 17 *S*s. Each two-alternative group received a different combination of two of the four patterns.

Procedure.—Each *S* had two sessions, a tachistoscopic session for threshold determination and a CRT session, spaced 1 wk. apart. The order of sessions was counterbalanced within groups so that about half the *S*s in each group received each order.

At the beginning of the first session, *S* was requested to describe in his own words each stimulus of the set on which he was to be tested. If his description contained the label intended by *E*, or an appropriate synonym, *E* selectively specified the label for the stimulus as "even," "three," "uneven," or "one-two." The *S* was instructed to use these labels in the experiment. A short practice run appropriate to the session was given in both tachistoscopic and CRT sessions. The *S* was rejected if he could not make an unambiguous description within an initial 30-sec. presentation of each stimulus appropriate to his group. Three *S*s were rejected for this reason. All stimulus presentations were preceded by a foreperiod of 1.0 sec. initiated by a .5-sec. ready signal (buzzer). The *S* fixated on a blank field prior to the exposure.

To determine thresholds, each pattern of the appropriate set was presented twice at each of 17 exposure durations ranging from .01 to 1.00 sec. The durations used were .01, .02, .03, .04, .06, .08, .10, .15, .20, .25, .30, .40, .50, .60, .70, .80, .90, 1.0 sec., with the 1.0-, .01-, and .03-sec. exposures omitted from the two-, three-, and four-alternative conditions, respectively. The patterns were presented by the method of constant stimuli with pattern and exposure-duration sequence randomized.

In the CRT session, each pattern was presented tachistoscopically 34 times on a random schedule. The *S* was instructed to depress a single microswitch key as soon as he could assign a standard label to the stimulus presentation. Following the key press which terminated the stimulus exposure, *S* was required to vocally report the label assigned to the exposure. Choice reaction times, measured from onset of the stimulus exposure to its termination by *S*, were recorded to the nearest .01 sec. A single key was utilized because of the possibility that the usual arrangement of separate keys for each response alternative introduced additional discriminative and perceptual-motor coordination factors.

Apparatus.—All stimuli were presented in a two-field Gerbrands mirror tachistoscope. Buzzer signal and foreperiod were controlled with Hunter timers. Choice reaction times were measured with a Hunter klockkounter.

Results and Discussion

A division of the 34 CRTs made by each *S* to each pattern into sequential thirds (after elimination of the longest CRT) indicated that CRTs became faster from the first third to the last third, and that the last 11 CRTs appeared consistently more stable than the earlier ones. All CRT results reported here were therefore based on a mean of the last 11 CRTs given by each *S* to each of the patterns to which he was required to respond.

The proportion of error responses to all stimuli in the CRT condition was very small. The mean proportions for two-, three-, and four-alternative conditions were, respectively, .01, .01, and .02. The range over the three conditions was .00–.08.

Table 1 shows the median of mean CRTs and median thresholds to each pattern in each combination (group). Analyses of variance indicated that in the two-alternative condition, differences in threshold between patterns within combinations exceeded the .05 significance level for the Ev-U, $F(1, 15) = 8.72$; the U-1,2, $F(1, 16) = 4.78$; and the 3-1,2 combinations, $F(1, 15) = 4.95$. In separate comparisons between CRTs within combinations in this condition, only the U-Ev combination reached significance, $F(1, 15) = 16.32$, $p < .01$. Analyses of variance of threshold differences between patterns in the

TABLE 1

Median Thresholds and Choice Times (CRT) for Each Pattern within Each Alternative Combination (Set)

	Stimuli within Two-Alternative Sets												
	Ev	U	Ev	3	Ev	1–2	U	3	U	1–2	3	1–2	$\bar{X}$
Mdn. CRT (msec.)	464	502	576	606	360	367	552	510	532	504	463	491	434
Mdn. threshold (msec.)	10	60	10	25	10	10	30	15	60	35	10	25	25

	Three-Alternative Set				Four-Alternative Set				
	Ev	1–2	3	$\bar{X}$	Ev	3	1–2	U	$\bar{X}$
Mdn. CRT (msec.)	576	614	643	611	685	856	910	1024	834
Mdn. threshold (msec.)	45	80	150	92	10	200	250	300	190

three- and four-alternative conditions all exceeded the .01 level of confidence, $F\ (2, 38) = 6.00$ and $F\ (3, 84) = 9.13$, respectively. Pattern CRTs differed significantly in the four-alternative, $F\ (3, 84) = 26.7, p < .001$, but not in the three-alternative condition.

The findings of major interest are the parallel between threshold and CRT. Table 1 shows that within each of the eight groups, the rank order of thresholds and CRTs was the same. Statistical tests (analyses of variance for the two-alternative condition and an order test developed by Page, 1963, for the other two conditions) evaluating the probability of obtaining such parallel orderings by chance in each combination, all led to the rejection of the null hypothesis beyond .001 level of confidence.

The obtained parallel between the pattern order for thresholds and CRT supports the assumption that at least part of the difference in the CRT of stimuli within each combination depends on differences in the amount of time required to receive sufficient information for the recognition of each pattern, i.e., the time which largely determines thresholds. This parallel may also be interpreted as indicating that the time represented by the threshold is not an artifact of the tachistoscopic method but reflects a general determinant of pattern perception.

Since thresholds are assumed to be a rough measure of the time required to obtain information about differences between spaces necessary for accurate percepts, it seems reasonable to suppose that most of the processing of such spacing information should begin only toward the end of the threshold period, since information which has not been received cannot be processed. Assuming constant processing time for stimuli within a combination, this formulation therefore predicts that CRT differences between alternative patterns should correspond to threshold differences. This expectation is closely approximated in several combinations shown in Table 1; E-U and 1,2-U in the two-alternative condition; E-1,2 in the three-alternative; and E-3, E-1,2, and 3-1,2 in the four-alternative conditions.

The above distinction between the temporal aspects of information reception and processing is entirely analytic, following logically from a formulation of the spatio-temporal determinants of information input (see also Kaswan et al., 1965). While this distinction implies some sort of neurophysiological information-assembly mechanism which combines information over time prior to processing, speculations about the characteristics of such a

system are beyond the scope of this paper.

Apart from the relatively brief, initial time component, the role of stimulus determinants of CRT is not clear. Thus, beyond the stimulus rank-order similarity for threshold and CRT within each stimulus combination, variations in CRT were, in many respects, different from threshold variations. For example, while threshold varied little for the Ev pattern in the different combinations, CRT for this pattern varied considerably. Similarly, Table 1 shows that rank order, from lowest to highest, of the two-alternative CRT means does not parallel that obtained for the two-alternative threshold means. For example, the longest CRT was for the 3 pattern in the Ev-3 combination, but the highest threshold was in response to the U pattern in the EV-U and 1,2-U combinations. Median CRT for patterns within each combination thus does not appear to be determined entirely by the difficulty of discriminating the design involved when threshold is the criterion for the ease of discrimination. This discussion indicates that unlike *relative* CRT to patterns *within* each of the eight combinations used, variations in CRT to each pattern in *different* combinations (e.g. CRT to the 1,2 pattern in the 1,2-Ev as compared to the 1,2-3 combination) may not be directly related to the time required to receive stimulus information.

REFERENCES

Kaswan, J., Young, S., & Nakamura, C. Y. Stimulus determinants of choice behavior in visual pattern discrimination. *J. exp. Psychol.*, 1965, **69**, 441–449.

Page, E. B. Ordered hypotheses for multiple treatments. A significance test for linear ranks. *J. Amer. Statist. Ass.*, 1963, **58**, 216–230.

(Received March 6, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 515–521

RESISTANCE TO EXTINCTION FOLLOWING BLOCKING OF THE INSTRUMENTAL RESPONSE DURING ACQUISITION[1]

W. EDWARD BACON

McGill University

4 groups of 11 rats each were trained to run for water in a right-turn alleyway and were then extinguished. Groups C_1 and C_2 received, respectively, 10 and 20 continuously reinforced training trials per day. Groups P and B received 20 trials daily on a random 50% reinforcement schedule. Group B Ss were prevented from running to the goal box on nonrewarded training trials by occlusion of the start box. Time measures indicated that the P and B groups were significantly more resistant to extinction than the C_1 and C_2 groups. Measures of irrelevant activities suggest that the blocking but not the partial-reinforcement variable alters start-box behavior.

It is established that a response is more resistant to extinction following partial reinforcement than following continuous reinforcement training (Lewis, 1960). There is also evidence, although inconclusive, that temporarily blocking the response at the goal box on some acquisition trials (partial delay of reinforcement) increases resistance to extinction (Renner, 1964). The literature, however, offers no information concerning the effect that permanently blocking the response in the alleyway on some acquisition trials has on extinction, although such blocking during extinction decreases resistance to extinction (Klugh, 1961; Lambert & Solomon, 1952). The aim of the present study was to provide such information. Following Bindra's (1963a) suggestion that "irrelevant" behavior occurring in the start box may be crucially related to extinction effects, various spontaneous activities of the animal in the start box were also recorded.

[1] This research was supported by a research grant (MH-03238-04) from the National Institutes of Mental Health to D. Bindra. The author is indebted to D. Bindra for his advice and support throughout the course of the experiment. Grateful acknowledgment is extended to L. Nadel for his assistance in collecting and analyzing the data.

METHOD

Subjects

The *Ss* were 44 male hooded rats, 71–88 days of age and weighing approximately 210 gm. at the start of the experiment. They were obtained from the Quebec Breeding Farm, Incorporated, St. Eustache, Province of Quebec, and were housed in standard colony cages, six per cage, throughout the study.

Apparatus

The apparatus was a right-turn, L-shaped black wooden alleyway, 9.5 in. high, covered with hinged Plexiglas sections. The main arm was 64 in. long and 4 in. wide, with the exception of the first 9 in. which was 6 in. wide and served as the start box. The short arm was 14 in. long and 6 in. wide and constituted the goal box. A black metal guillotine door separated the start box from the running alley and a similar door located 39.5 in. down the running alley marked the beginning of the goal box. A photoelectric cell, activated by a light beam passing through an infrared filter, was situated in the side wall of the running alley 1.5 in. above the floor and 2.5 in. from the start-box door. A second photoelectric cell, activated by a similar light source, was located down the running alley 42 in. from the first photoelectric cell. A metal drinking tube containing a $\frac{3}{16}$-in. fluid hole projected through a $\frac{3}{4}$-in. hole cut in the far wall of the goal box 3 in. above the floor. On nonreinforced trials a 6×9.5 in. black

false wall of $\frac{1}{8}$-in. plywood covered the hole in the goal-box wall. On "blocked response trials" a 4 × 9.5 in. black wooden partition was inserted into the running alley approximately 1 in. from the start-box door.

The raising of the start-box door started a .01-sec. Standard Electric timer. This timer stopped and a second similar timer started when *S* interrupted the first light beam. The second timer stopped when *S* passed through the second light beam. The timers registered starting and running times, respectively.

Procedure

Taming and habituation.—Ten days prior to pretraining, *S*s were placed on a 20–23.5 hr. water-deprivation schedule, which consisted of $\frac{1}{2}$ hr. access to water daily and which was maintained throughout the experiment. Each *S* was given 2–4 min. of gentling daily on at least 5 of the 10 days prior to pretraining. On the fourth and third days prior to pretraining, thirsty *S*s were placed in the alleyway goal box (goal-box door closed) for 5 min. with water present.

Pretraining.—Each *S* was given five pretraining trials. A trial consisted of placing *S* in the start box, raising the start-box door after a 15-sec. delay, lowering the start- and goal-box doors after *S* passed under them, and allowing *S* 15 sec. access to water in the goal box. The intertrial interval was approximately 45 sec. and each rat was given $\frac{1}{2}$ hr. access to water approximately $\frac{1}{2}$ hr. after the last trial of the day.

Measures of each *S*'s activity were obtained during the 15-sec. period of delay in the start box by means of a modified time-sample method (Bindra & Blond, 1958). On each trial, five samples of a start-box activity were recorded. At the end of each 3-sec. interval, the "on-going" activity was categorized and recorded as either grooming (G), sitting (L), sitting-sniffing (L-S), rearing (R), rearing-sniffing (R-S), or locomotion (W). These categories have previously been defined (Bindra, 1963a; Bindra & Blond, 1958).

Acquisition training.—Following pretraining, *S*s were randomly assigned to four groups, 11 per group, for 6 days of acquisition training. Group P received 20 trials per day on a 50% reinforcement schedule randomized within blocks of 20 trials with the restrictions that the first and the last trials of each block be rewarded and that runs longer than 3 rewarded or 3 nonrewarded trials be excluded. Group B received 20 trials per day with 50% of the trials being blocked response trials. On these trials the start-box door was opened after the 15-sec. delay but the inserted partition prevented entry into the running alley. The *S* was removed from the start box 15 sec. after the door was opened. The ordinal position of the blocked trials corresponded to nonrewarded trials for Group P. Groups C_1 and C_2 received 10 and 20 trials per day, respectively, on a 100% reinforcement schedule. These two continuously reinforced groups were included in the experimental design as Group B *S*s received only 10 opportunities to run to the goal box out of 20 daily trials. All other running and recording procedures were similar to those employed during pretraining with the exception that start-box activity was recorded only on the first 5 and last 5 trials of each day.

Extinction.—Extinction trials began the day following the completion of acquisition training and continued for 3 days. All *S*s received 20 nonreward trials per day provided they did not meet the extinction criterion of failing to leave the start box within 2 min. or to enter the goal box 2 min. after leaving the start box on three successive daily trials. Extinction trials were terminated for that day when *S* fulfilled the criterion, and start- and running-time scores of 120 sec. were assigned to the uncompleted trials. Other running and recording procedures were similar to acquisition training.

Results

Start-Box Activity

The frequency of occurrence of each category of start-box activity was determined for each *S* for the pretraining trials and for each daily block of training and extinction trials. The daily frequency measures were converted to proportions by dividing the frequency measure by the total number of daily observations on each *S*. The proportions were transformed to arc-sine scores as suggested by Mosteller and Bush (1954) and the arc-sine scores for each of the six categories of start-box activity were subjected to separate analyses of variance.

The results of the analyses (Table 1) indicate that reliable group differences occurred only for the activities of R-S during acquisition and R-S and L-S

TABLE 1

ANALYSIS OF VARIANCE ON THE ARC-SINE TRANSFORMATIONS OF FREQUENCY OF OCCURRENCE OF START-BOX ACTIVITIES DURING PRETRAINING, ACQUISITION, AND EXTINCTION

Source	*df*	*F*					
		L	R	G	L-S	R-S	W
Pretraining							
Between							
Groups (G)	3	0.68	0.72	2.11	0.16	0.25	1.42
Error (*MS*)	40	(156.8)	(66.5)	(44.7)	(68.1)	(48.3)	(53.1)
Acquisition							
Between							
G	3	2.35	1.01	1.99	2.30	4.29*	0.52
Error (*MS*)	40	(259.8)	(101.1)	(230.3)	(235.8)	(269.5)	(166.1)
Within							
Days (D)	5	10.91**	0.90	4.06**	1.45	2.49*	0.84
G × D	15	1.17	1.39	1.18	1.63	0.54	0.46
Error (*MS*)	200	(43.3)	(20.0)	(29.1)	(38.2)	(39.4)	(28.0)
Extinction							
Between							
G	3	0.75	0.97	1.01	3.20*	4.37**	0.16
Error (*MS*)	40	(298.5)	(16.5)	(274.7)	(164.0)	(195.3)	(89.6)
Within							
D	2	2.28	2.93	8.97**	1.63	4.39**	2.28
G × D	6	0.43	1.64	0.73	1.76	1.23	0.77
Error (*MS*)	80	(95.2)	(9.9)	(43.5)	(45.2)	(44.2)	(32.1)

* $p < .05$.
** $p < .01$.

during extinction. Scheffe's test for multiple comparisons (Edwards, 1960, pp. 154–156) on the overall mean acquisition R-S scores indicates that Group B *S*s did significantly less rearing-sniffing than Group C_2 *S*s ($p < .05$). Other group comparisons were not statistically reliable (p's $> .05$). Similar comparisons carried out separately on the overall mean extinction R-S and L-S scores revealed that *S*s in Group B did significantly less rearing-sniffing than *S*s in the other three groups (p's $< .05$) and significantly more sitting-sniffing than Group P *S*s ($p < .05$). All other comparisons proved to be nonsignificant (p's $> .05$).

Three start-box activities showed reliable changes across trials (Day effects). There was a tendency for R-S to increase as a function of acquisition trials and decrease as a function of extinction trials. The frequency of occurrence of L steadily declined during acquisition but was not reliably affected by extinction. Grooming increased during early acquisition and then showed a further increase during extinction.

Instrumental Response

Pretraining.—Start and running times were averaged separately across the five pretraining trials for each *S*. Analyses of variance on the mean time scores indicated no statistically significant differences between the four groups for either start time, $F(3, 40) = 0.04$, $p > .05$, or running time, $F(3, 40) = 1.46$, $p > 0.5$.

Acquisition training.—Mean start-time and running-time scores were computed for each daily block of trials

TABLE 2

MEAN RUNNING TIME VARIABILITY SCORES FOR THE FOUR EXPERIMENTAL GROUPS DURING EARLY AND LATE ACQUISITION

Groups	Early Acquisition	Late Acquisition
C_1	10.80	9.77
C_2	28.48	6.59
B	15.25	4.57
P	35.69	2.97

(20 trials per block for Groups C_2 and P and 10 trials per block for Groups C_1 and B). Analyses of variance indicated a reliable Trial Block effect for both start time, $F\ (5, 200) = 15.25$, $p < .01$, and running-time means, $F\ (5, 200) = 15.87$, $p < .01$. However, the Group effect was not significant for either start time, $F\ (3, 40) = 1.21$, $p > .05$, or running time, $F\ (3, 40) = 0.81$, $p > .05$; nor was the Group $\times$ Trial Block interaction significant for start time, $F\ (15, 200) = 1.10$, $p > .05$, or running time, $F\ (15, 200) = 0.80$, $p > .05$. In general, although performance of the four groups improved with trials, there were no differences in performance between the four groups when response times were averaged across blocks of daily trials.

To determine whether the variables of nonreward and blocking increased response variability, the ranges of the start times and running times for the first 20 (early acquisition) and last 20 acquisition trials (late acquisition) were determined for blocks of 5 trials and averaged to yield early and late acquisition variability scores. Analyses of variance revealed that animals were considerably less variable during late as compared to early acquisition for both start-time, $F\ (1, 40) = 20.16$, $p < .01$, and running-time measures, $F\ (1, 40) = 18.71$, $p < .01$. The only indication of a difference in variability between the four groups was the significant Group $\times$ Early-Late Acquisition interaction (Table 2) for the running-time measure, $F(3,40) = 3.21$, $p < .05$. Scheffe's test for multiple comparisons showed no reliable differences (p's $> .10$) between any of the groups during late acquisition. Only the P vs. C_1 difference was significant ($p < .05$) among the early acquisition comparisons of pairs of means. Examination of the group means in Table 2 and the significant $C_2 + P$ vs. $C_1 + B$ comparison ($p < .05$) suggests that response variability is greater during early acquisition for the two groups receiving 20 daily trials (C_2 and P).

Extinction.—Start times and running times were averaged across blocks of 20 extinction trials for each *S* and the trial-block means were subjected to analyses of variance. The results of the analysis on start time showed a statistically reliable Group effect, $F\ (3, 40) = 4.32$, $p < .01$, and Trial Block effect, $F\ (2, 80) = 10.09, p < .01$, but a nonsignificant Group $\times$ Trial Block interaction, $F\ (6, 80) = 0.77$, $p > .05$. A similar analysis yielded a significant Group effect, $F\ (3, 40) = 6.43$, $p < .01$, and Trial Block effect, $F\ (2, 80) = 13.34$, $p < .01$, but a nonsignificant interaction, $F\ (6, 80) = 1.04$, $p > .05$, for running time. Figure 1 presents mean response times over blocks of 20 trials for the four groups. Both start- and running-time measures increased across trial blocks for all groups. Examination of the individual group performance curves reveals that there was little difference between Groups C_1 and C_2 or between Groups B and P. Scheffe's test indicated that these comparisons were not statistically reliable for either the overall start- or running-time means (p's $> .10$). However, the $C_1 + C_2$ vs. $B + P$ comparisons were significant for both time measures (p's $< .05$).

The blocked and partially reinforced *Ss* showed greater resistance to extinction than the continuously reinforced *Ss*.

Visual inspection of the group performance curves also suggests that Group B *Ss* showed somewhat less initial resistance to extinction but better terminal extinction performance than Group P *Ss*. Analyses of variance were carried out on five-trial start- and running-time means for the Group B and P data only. Although the Group × Trial Block interaction was not reliable for start time, $F\ (11, 220) = 1.07$, $p > .05$, the sig-

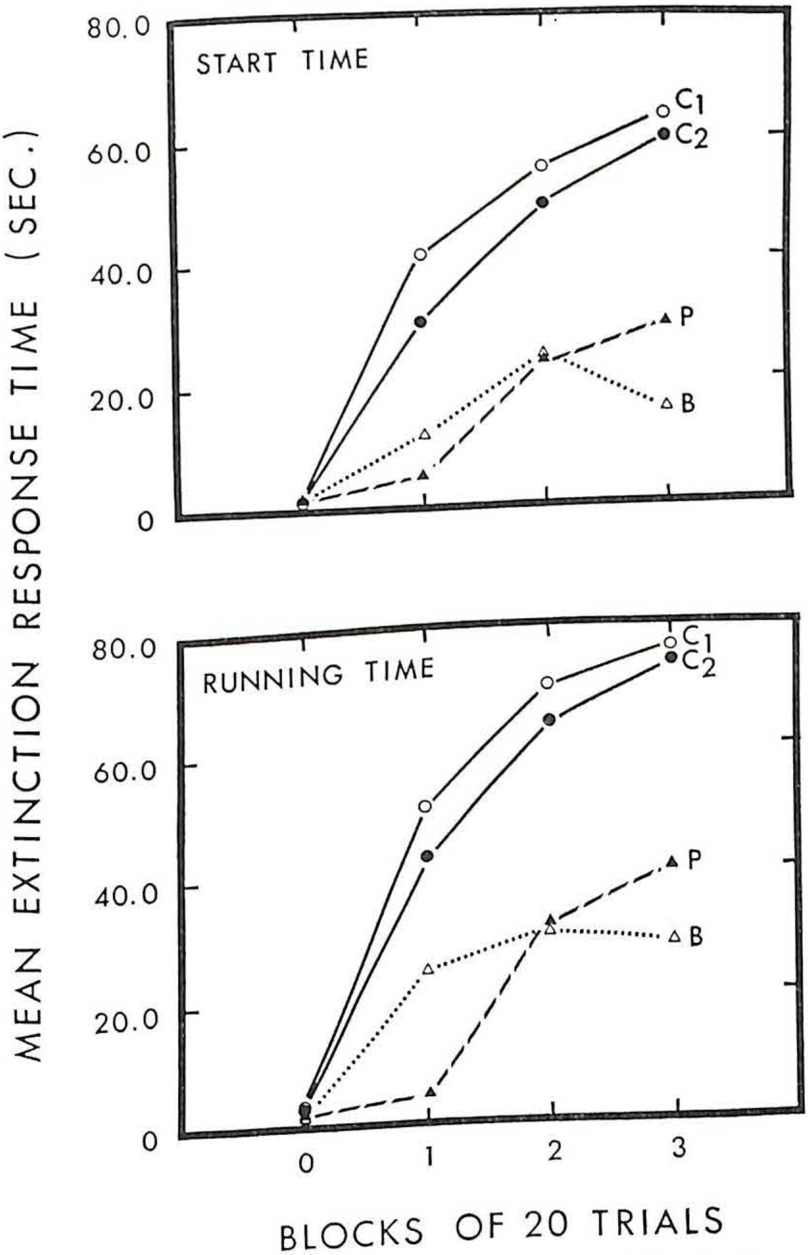


FIG. 1. Mean start and running time across blocks of trials for the four experimental groups. (Block 0 indicates the mean start or running time on the last day of acquisition and Blocks 1–3 the 3 extinction days.)

nificant Group × Trial Block interaction for running time, $F(11, 220) = 2.78$, $p < .05$, provides some support for the above suggestion.

Discussion

The changes in occurrence of grooming, sitting, and rearing-sniffing in the start box across acquisition and extinction trials agree with data reported by Bindra (1963a) with one exception. Bindra found that the duration of time spent sitting in the start box decreased across extinction trials; whereas, in the present experiment, the frequency of occurrence of sitting did not show any reliable change during extinction. This discrepancy in findings may be due to the different methods of measuring start-box activity used in the two studies.

Bindra (1963b) has presented evidence suggesting that in a lever-pressing situation irrelevant activities are related to such phenomena as the partial-reinforcement effect. He contends that in an alleyway situation the effect of goal-box events such as nonreward is to alter subsequent behavior in the start box which in turn affects the instrumental response (Bindra, 1963a). The absence of any difference in start-box activities between partially and continuously reinforced *S*s, together with the presence of a clear difference in instrumental response times between these groups during extinction, does not support Bindra's contention. However, the finding that blocking did alter start-box behavior, particularly rearing-sniffing, suggests that the relation of start-box activity to extinction phenomena is in need of further clarification.

Of special theoretical interest is the finding that the blocked response condition resulted in a marked increase in resistance to extinction; almost, if not as great, as that due to partial reinforcement. Lawrence and Festinger (1962) propose that in an alleyway situation certain events such as nonreward cause dissonance in the rat. On nonrewarded acquisition trials partially reinforced *S*s develop "extra attractions" in the goal box which reduce dissonance and, in turn, increase resistance to extinction. Blocked *S*s never experienced nonreward in the goal box during acquisition and hence should not have developed any extra goal-box preferences. The increased resistance to extinction following blocking of the instrumental response would not seem compatible with dissonance theory as developed so far.

On the other hand, the extinction findings can be accounted for within the framework of frustrative-nonreward theory of Amsel (1958, 1962) if it is assumed that blocking, like nonreward, elicits frustration and, further, that a secondary form of frustration similar to the fractional anticipatory frustration reaction (r_f) can work *forward* in the alleyway through stimulus generalization. If this be the case, both blocked and partially reinforced *S*s would have become conditioned to run to cues of frustration (s_f) associated with r_f. It should be noted that P and B animals did not exhibit significantly greater response variability during early acquisition than respective C *S*s with an equal number of daily trials as predicted by frustration theory. Subsequent analysis suggests that increased variability due to frustration may have been obscured by factors associated with closely massed trials.

The present study yields at least three kinds of information. First, it provides empirical evidence that the blocking of an instrumental response during acquisition increases its resistance to extinction. Secondly, it casts considerable doubt upon the utility of the cognitive-dissonance interpretation of extinction effects. Thirdly, it suggests that, in arriving at an explanation of extinction phenomena, one should not limit oneself only to events occurring in the goal box. Events occurring in other parts of the alleyway may be equally important.

References

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

Amsel, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **69**, 306–328.

Bindra, D. Effects of nonreinforcement, drive-diminution, and punishment on relevant and irrelevant behavior in the runway. *Psychol. Rec.*, 1963, **13**, 272–283. (a)

Bindra, D. Temporal analysis of relevant and irrelevant behavior components in partial reinforcement and extinction. *Psychol. Rep.*, 1963, **13**, 551–563. (b)

Bindra, D., & Blond, J. A time-sample method for measuring general activity and its components. *Canad. J. Psychol.*, 1958, **12**, 74–76.

Edwards, A. *Experimental design in psychological research.* (Rev. ed.) New York: Rinehart, 1960.

Klugh, H. E. Speed of running in extinction as a function of differential goal box retention time. *J. exp. Psychol.*, 1961, **61**, 172–177.

Lambert, W. W., & Solomon, R. L. Extinction of a running response as a function of distance of block point from the goal. *J. comp. physiol. Psychol.*, 1952, **45**, 269–279.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcement.* Stanford: Stanford Univer. Press, 1962.

Lewis, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–28.

Mosteller, F., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954. Pp. 289–334.

Renner, K. E. Delay of reinforcement: A historical review. *Psychol. Bull.*, 1964, **61**, 341–361.

(Received March 7, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 522–527

SAMPLE SIZE AND THE REVISION OF SUBJECTIVE PROBABILITIES [1]

CAMERON R. PETERSON,[2] ROBERT J. SCHNEIDER, AND ALAN J. MILLER

University of Colorado

This experiment investigates the relation between subjective probability and mathematical probability. Bayes' theorem provides the correct revision of probabilities as a result of new data. New data, however, may come in different sized samples. This experiment measured the accuracy of subjective probability revision as a function of sample size. The results show that accuracy decreases as sample size increases; the function is negatively accelerated. The gain in amount of data processed by increasing the sample size is at the expense of accuracy.

Behavioral decision theory provides a framework for evaluating the extent to which selected behavior of human *S*s corresponds with ideally consistent behavior as outlined by statistical decision theory (Edwards, 1961). The two primary variables of decision theory are probability and value; the corresponding psychological variables are subjective probability and utility. The present experiment studies the correspondence between subjective probability and mathematical probability.

Recent developments within statistical decision theory focus on the correct revision of probabilities in the light of new information—the problem of revising the probability of a hypothesis as a function of the occurrence of a relevant datum (e.g., see Schlaifer, 1961; or Raiffa & Schlaifer, 1961). These developments revolve around the concepts of personal probability and Bayes' theorem. Personal probabilities are ideally consistent opinions, and conform to the axioms of probability theory; subjective probabilities are the opinions of real persons, and may or may not conform to the axioms of probability theory. Bayes' theorem is a derivation from the axioms of probability theory. A consequence of Bayes' theorem which is relevant to this paper is

$$\frac{P(H_a|D)}{P(H_b|D)} = \frac{P(D|H_a)\,P(H_a)}{P(D|H_b)\,P(H_b)} \qquad [1]$$

or, more simply,

$$\Omega_1 = L\Omega_0 \qquad [2]$$

where H_a and H_b refer to two different hypotheses regarding the state of the world, and D refers to a relevant datum which occurs. The Ω_0 refers to the odds of H_a to H_b prior to the occurrence of D. The Ω_1 refers to the corresponding odds posterior to the occurrence of D. The L refers to a likelihood ratio; it is equal to the conditional probability of D given H_a divided by the conditional probability of D given H_b. Thus Equation 2 specifies that the revision of the odds of H_a to H_b is a function of the occur-

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 42 of the Institute. A more detailed report of this experiment, including additional analyses, is available as Behavior Research Laboratory Report No. 41 (mimeo). The research was supported by a research grant (M-4977) from the National Institute of Mental Health. The authors are indebted to W. Edwards for his continual guidance. The many suggestions of K. R. Hammond and F. J. Todd and the assistance of Dorothy Boucher in the conduct of the experiment are gratefully acknowledged.

[2] Now at the University of Michigan.

rence of D. This revision is accomplished by multiplying the prior odds by the likelihood ratio to yield the posterior odds (for a detailed discussion of Equation 2 see Edwards, Lindman, & Savage, 1963, p. 218).

Experiments carried out by Edwards and his associates indicate that *S*s revise their subjective probabilities in the correct direction, but that the magnitude of revision is less than that specified by Bayes' theorem (e.g., Edwards & Phillips, 1963, 1964). Typically, in these experiments, *S*'s task has been to revise his subjective probability about which H was generating a sequence of data. A revision was made after the presentation of each datum in the sequence.

Information which results in a revision of subjective probabilities need not be restricted to a single datum at a time. Information may be provided in a sample of data. But more data in a sample implies more information processing. For some types of human information processing, an increase in the amount of information presented to an *S* impairs his accuracy; for other types of information processing, the chunking of more information into a presentation does not impair accuracy (Miller, 1956). The primary purpose of the present experiment was to investigate the function relating accuracy of data processing to the *size* of the *sample.* Accuracy was measured by the degree to which *S*s' revision of subjective probabilities agreed with the correct revision. Sample size referred to the specific number of data in the sample which provided the information for the revision of subjective probabilities.

Method

Experimental design.—The experimental design required the manipulation of sample size and an evaluation of the effect upon accuracy in the revision of subjective probabilities. Urns with varying proportions of black marbles and white marbles were the *H*s. Each trial consisted of a random selection of 48 marbles (*D*s) from an urn. After each trial *S*s were shown which urn yielded the marbles.

Subjects.—Forty-four volunteer students from an introductory psychology course served as *S*s in groups of four.

Apparatus.—Two large beakers served as urns. Each urn contained approximately 1,000 marbles of $\frac{1}{2}$ in. diameter. Large numbers of marbles were used so that, with continual mixing, sampling without replacement could be considered a reasonable approximation of sampling with replacement. Urn B contained three black marbles for every two white marbles; Urn W contained two black marbles for every three white marbles. In this symmetrical situation the ratio of most probable to least probable color was a constant 60:40 for both urns. This relatively uncertain ratio insured that *S*s were not required to make too many revisions near a ceiling of certainty, i.e., subjective probabilities of 1 or 0.

Four sample sizes were selected: 1, 4, 12, and 48. Relatively more small sizes were included because of the expectation that the function would be negatively accelerated. For samples of size 1, *E* drew single marbles at a time. Dippers were made which would "dip" the appropriate number of marbles for each sample.

The response apparatus for obtaining subjective probability estimates was a 25-in. bar marked off into 100 equal units which were numbered from left to right. A sliding marker permitted *S* to divide the bar into two sections. The length of the left section represented *S*'s subjective probability that the marbles observed were drawn from Urn B; the length of the right section represented *S*'s subjective probability that the marbles were drawn from Urn W. The length of the entire rod represented a probability of 1, so subjective probabilities for both urns were required to sum to 1. The number under the sliding marker represented the subjective probability of Urn B. In order to preclude problems inherent in responses of complete certainty, subjective probabilities were coded as no more extreme than .001 or .999.

Black drapes divided one side of the experimental room into four cubicles, each with an opening facing *E*. This permitted four individual *S*s to be run simultaneously; but the experimental atmosphere approximated

the running of individual *S*s in terms of both communication between *S*s and rapport with *E*.

Procedure.—The experiment consisted of eight blocks of four trials each. Each trial consisted of one of the four sample sizes. During each trial *E* drew and displayed successive samples of marbles until the cumulative number of marbles drawn numbered 48. After each sample *S* was required to revise and record his subjective probability that *E* was drawing from Urn B. In order to avoid confounding the effects of sample size and within-block learning, the sample sizes were counterbalanced with respect to order of occurrence during a block.

The *S*s, four or fewer, were seated in cubicles facing *E*. A table in front of each *S* contained a probability bar and an answer sheet. The *E* instructed *S*s with respect to the content of the urns and the use of the probability bar. For example, if *S* expected that it was twice as likely that *E* was drawing from Urn B rather than from Urn W, he would set the marker on the probability bar so that the Urn B side was twice the length of the Urn W side. The *S*s were told that the purpose of the experiment was to compare their intuitive revision of subjective probabilities with the correct revision as calculated by means of a mathematical equation. There was a practice demonstration of the step-by-step procedure.

The procedure of each trial progressed according to the following steps. (*a*) The *E* flipped a coin and selected the designated urn. (*b*) The *E* placed the urn into an opaque box so that *S*s were not able to identify the urn. (*c*) Each *S* set the marker on the probability bar at .50. (*d*) The *E* drew at random a prescribed sample size of marbles and displayed the sample to *S*s. (*e*) On the basis of the sample, each *S* reset his probability marker and recorded his subjective probability of Urn B on his answer sheet. This sequence of draw-reset marker-record was repeated until *E* had drawn 48 marbles from the urn. (*f*) The *E* then showed *S*s which urn he had been drawing from. He then progressed to the next trial by flipping another coin to select the next urn. This procedure continued until the end of the experiment.

Measure of probability revision.—In order to compare the subjective probability revision of *S*s with the corresponding probability revision as calculated by means of Bayes' theorem, the data analyses required a common measure of revision. The measure used here was the same as that reported by Edwards and Phillips (1963). It is the $\log_{10}$ of the likelihood ratio (LLR). The likelihood ratio is described above in Equation 2; it is equal to the ratio of the posterior to the prior odds. Thus, the LLR of this experiment was equal to the log of the posterior odds in favor of Urn B minus the log of the corresponding prior odds. Since the LLR increases with the difference between the posterior and prior odds, it must also increase with the difference between the posterior and prior probabilities. It is thus a measure of revision from prior to posterior subjective probabilities; the measure is positive if the revision is upward and negative if the revision is downward.

The LLRs for the present experiment were calculated as follows. Posterior probabilities, both subjective and Bayesian, were obtained after each trial of 48 marbles. These posterior probabilities were converted into odds in favor of Urn B, and were then transformed into log posterior odds. The definition of LLR calls for a subtraction of log prior odds from log posterior odds. However, since the initial probability of Urn B on each trial was .50, the prior odds were always 1 and the log prior odds were 0. Therefore, the log posterior odds were equivalent to LLRs. An LLR calculated from Bayesian probabilities will be termed BLLR: one calculated from subjective probabilities will be termed SLLR. The BLLR in the present experiment was equal to .17609 times the total number of black marbles minus white marbles.

An important advantage of the LLR over the algebraic change in probabilities as a measure of probability revision is that the LLR is proportional to the amount of evidence in favor of one hypothesis over the other. Algebraic change in probabilities, on the other hand, is influenced by a ceiling of certainty; equal increments of evidence result in smaller expected algebraic probability changes near certainty than near uncertainty.

Relation of SLLR to BLLR.—Two dependent variables related SLLRs to corresponding BLLRs. The primary dependent variable, the *accuracy ratio*, was the ratio of SLLR to BLLR at the end of each trial of 48 marbles. If *S* changed his probabilities in the same algebraic magnitude as did Bayes' theorem, then his accuracy ratio would be 1. To the extent that *S* changed less than Bayes' theorem, his accuracy ratio decreased.

A second dependent variable, linearity, was the correlation between corresponding within-trial BLLRs and SLLRs. The LLRs were based on probabilities posterior to each sample, i.e., posterior to each probability

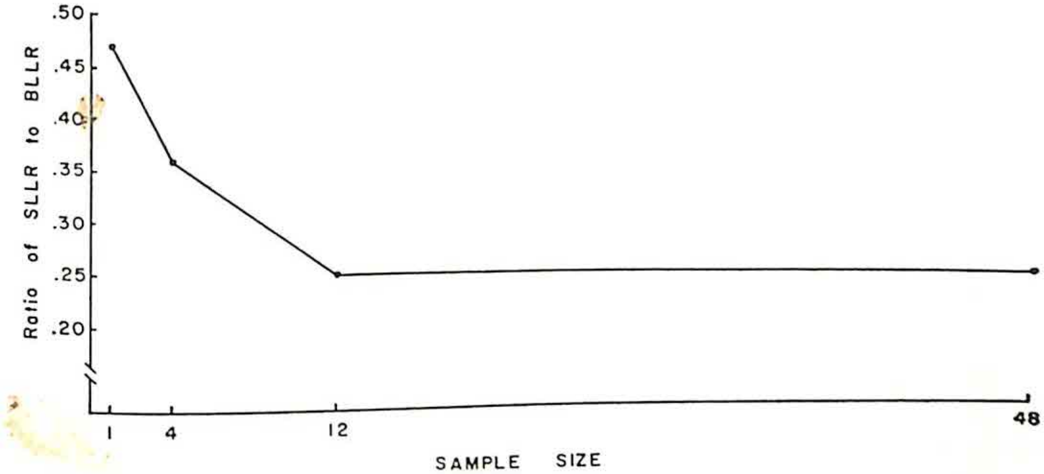


FIG. 1. Mean accuracy ratio as a function of sample size.

revision. A correlation was calculated for each trial of sample sizes 1 and 4; there were too few within-trial LLRs for sample sizes of 12 and 48.

RESULTS

Analyses of variance.—In order to evaluate the significance of the effect of sample size upon the accuracy of subjective probability revision, analyses of variance were applied to the accuracy ratio and to linearity. The two main effects for these analyses were sample size and blocks of trials. These analyses of variance are presented in Table 1. Whenever the number of black and white marbles was exactly equal, BLLR had a value of 0 and thus the accuracy ratio was indeterminate. As a constantly conservative procedure, each indeterminate value was replaced by the grand mean of the data in the analysis of variance and the degrees of freedom associated with the highest order error term were reduced by one.

Table 1 indicates that sample size had a significant ($p < .001$) effect upon the accuracy ratio. No other F ratio reached the .01 level of significance.

Description of effects.—Table 2 presents the mean value of each dependent variable associated with each level of the independent variables. Figure 1 illustrates the function relating accuracy ratio to sample size. This ratio decreases rapidly at first, and then more slowly, as sample size increases.

Although the LLR is the more adequate measure of probability revision, Fig. 2 further illustrates the effect of sample size by a more direct measure—absolute probability change. The horizontal axis indicates the absolute difference between black and white marbles at the end of each trial. The vertical axis indicates the net absolute revision from the prior probability to the posterior probability at the end of

TABLE 1

ANALYSES OF VARIANCE ON ACCURACY RATIO AND LINEARITY

Source	Accuracy Ratio		Linearity	
	df	*F*	*df*	*F*
Sample Size (A)	3/129	13.72***	1/43	4.04
Blocks (B)	7/301	1.32	7/301	2.57*
A × B	21/903	1.78	7/301	<1

* $p < .05$.
*** $p < .001$.

TABLE 2

MEAN VALUE FOR EACH DEPENDENT VARIABLE FOR EACH LEVEL OF THE INDEPENDENT VARIABLES

	Accuracy Ratio	Linearity
Sample Size 1	.470	.915
Sample Size 4	.361	.947
Sample Size 12	.253	—
Sample Size 48	.255	—
Block 1	.349	.929
Block 2	.342	.910
Block 3	.327	.951
Block 4	.305	.917
Block 5	.283	.943
Block 6	.328	.917
Block 7	.363	.940
Block 8	.384	.944

the trial. The solid curve represents the correct revision as calculated by means of Bayes' theorem. The three dotted lines are smooth curves intended to represent the data points associated with sample sizes of (*a*) 1, (*b*) 4, and (*c*) 12 and 48. The observation base varied across data points, and was relatively small when abscissa values were more extreme than 2 and 22. Figures 1 and 2 tell the same story. Subjective probability revision lagged behind Bayesian revision; amount of revision decreased from sample size 1 to 4, decreased from 4 to 12, and remained constant from 12 to 48.

Means and individual differences.—The first two rows of Table 3 present the mean and the standard error of the mean for the accuracy ratio and linearity. In order to evaluate the consistency of individual differences across situations, scores were compared among all levels of sample size. By collapsing over blocks, one score was obtained for each *S* in each sample size. Correlating each pair of sample sizes across *S*s resulted in a matrix of

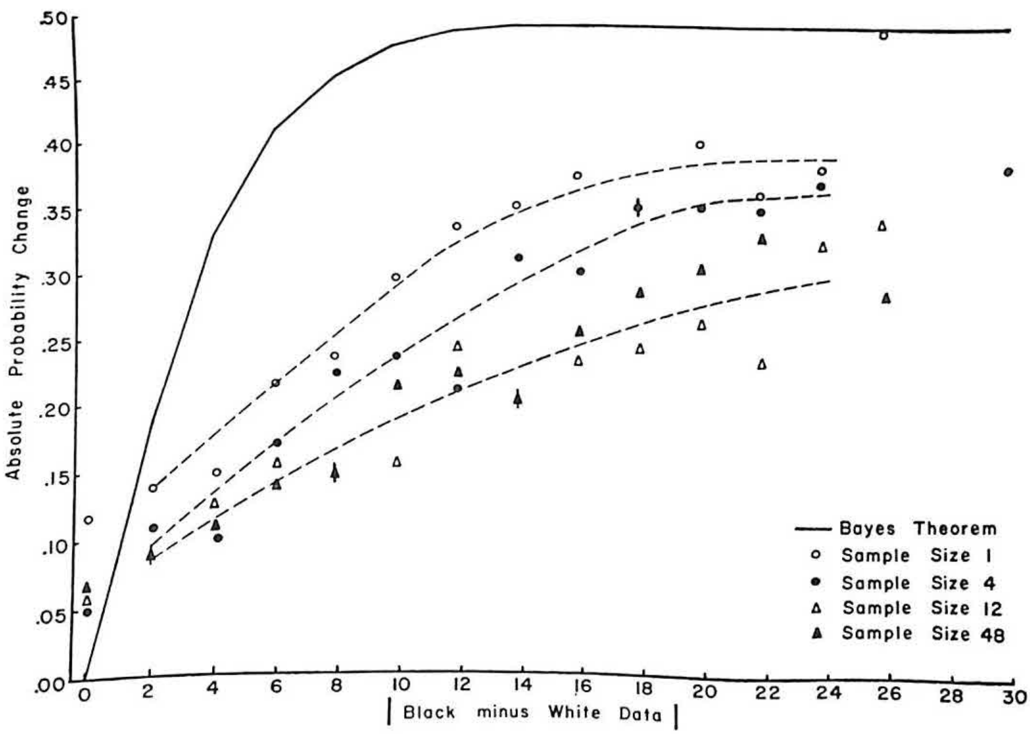


FIG. 2. Net absolute probability change as a function of the absolute difference between black and white data at the end of a trial. (A verticle line through a solid data point indicates coincidence of open and closed points.)

six correlations. The average of these six correlations is presented in the bottom row of Table 3, indicating the extent to which an *S*'s performance in one sample size could be predicted by his performance in another sample size. With $df = 42$, a correlation of .38 is significantly different ($p < .01$) from zero.

TABLE 3

MEAN, VARIABILITY, AND CONSISTENCY MEASURES ACROSS *S*s FOR THE ACCURACY RATIO AND LINEARITY

	Accuracy Ratio	Linearity
M	.334	.933
σ_M	.039	.009
Consistency	.675	.518

DISCUSSION

The primary purpose of the present experiment was to investigate the shape of the function which relates accuracy in the revision of subjective probabilities to the size of the sample of data upon which the revision is based. The graphs of Fig. 1 and 2 answer this question. Accuracy of subjective probability revision decreases with a negatively accelerated slope as sample size increases.

With reference to the information processing findings discussed earlier, these results provide another instance in which the accuracy of human information transmission decreases as the amount of information to be processed per observation increases. However, even though accuracy decreases, an increase in sample size results in more data processing per observation. Changing from sample size 48 to sample size 1 in the present experiment resulted only in a doubling of accuracy at the cost of 48 times as many observations. Thus, greater revision per observation results from increasing sample size, but at a cost in accuracy.

The lack of effect of any variable upon linearity means that *S*s in the present experiment were equally linear with Bayes' theorem, regardless of the level of sample size or block. Of course, caution is to be used in the interpretation of this meaning of linearity. Any variance in subjective probability revision which was not accounted for by Bayes' theorem may be due either to a systematic nonlinear relation or to error variance around a straight line. However, the relatively consistent magnitude of correlations reported in Table 2 (all mean correlations in the table fall between .91 and .95) implies an important degree of linearity between Bayes' theorem and the revision of subjective probabilities.

Finally, the experimental results provide further support for the previous finding that the revision of subjective probabilities lags behind the corresponding optimal revision as calculated via Bayes' theorem. Fairly high linearity seems to characterize this relation. Some *S*s lag much more than others; the degree of lag is significantly consistent from one sample size to another. Some *S*s are more Bayesian than others.

REFERENCES

EDWARDS, W. Behavioral decision theory. *Ann. Rev. Psychol.*, 1961, **12**, 473–498.

EDWARDS, W., LINDMAN, H., & SAVAGE, L. J. Bayesian statistical inference for psychological research. *Psychol. Rev.*, 1963, **70**, 193–242.

EDWARDS, W., & PHILLIPS, L. D. Conservatism in probabilistic inference. Paper read at Midwestern Psychological Association meetings, Chicago, 1963.

EDWARDS, W., & PHILLIPS, L. D. Man as a transducer for probabilities in Bayesian command and control systems. In G. L. Bryan & M. W. Shelly (Eds.), *Human judgments and optimality*. New York: Wiley, 1964.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

RAIFFA, H., & SCHLAIFER, R. *Applied statistical decision theory*. Boston: Harvard University, Graduate School of Business Administration, Division of Research, 1961.

SCHLAIFER, R. *Introduction to statistics for business decisions*. New York: McGraw-Hill, 1961.

(Received March 12, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 528–533

EFFECTS OF POSTRESPONSE STIMULUS DURATION UPON DISCRIMINATION LEARNING IN HUMAN SUBJECTS [1]

DONALD J. DICKERSON AND NORMAN R. ELLIS

George Peabody College for Teachers

3 experiments, designed to explore temporal relations in instrumental conditioning, are reported. The variable manipulated was postresponse stimulus duration (PSD), the length of the interval from S's response to the offset of the discriminanda. The tasks were 2-choice discrimination problems. In a 1st experiment PSDs were 0.0, 0.5, 1.0, and 2.0 sec. Normal and retarded human Ss were assigned to each PSD group. A 2nd experiment was similar to the 1st except the duration of the reinforcing light was changed from 1.0 sec. to 3.0 sec. A 3rd experiment differed in several ways from the 1st 2. The cues for the task were different, normal Ss only were used, a buzzer instead of the light served as a reinforcer, and the PSDs were 0.0, 0.2, 0.5, 0.75, and 1.0 sec. The main findings show that PSD has a marked effect upon discrimination learning of the type studied. The optimal interval was 0.5 sec. PSDs of shorter and longer durations led to slower learning. Normal Ss learned faster than retarded Ss, although no interaction effects prevailed.

Temporal relations between events in classical conditioning have received extensive study, and the main findings have been summarized by Kimble (1961). On the other hand, temporal aspects of instrumental conditioning have received little experimental attention. The most effort has focused on delay of reinforcement. Although numerous investigators have dealt with comparisons of simultaneous and successive discrimination learning, such studies are complex and appear to involve variation along other dimensions.

Recent experimental evidence indicates that variation of temporal relations between stimuli, responses, and reinforcements has marked effects upon discrimination learning. Chow and Orbach (1957) and Riopelle and Addison (1962) have studied the effect of stimulus duration upon the rate of simultaneous discrimination learning in rhesus monkeys. In the main, these studies show that learning is facilitated by longer exposure to the discriminanda. Learning rate increases as stimulus duration increases from 0.1 sec., reaching an asymptote at around 1.0 sec. Sheridan, Horel, and Meyer (1962) report that the postresponse persistence of either the positive or negative cue produces learning superior to that observed when both or neither of the cues persist.

The present experiments assess the effects of postresponse stimulus duration (PSD) upon two-choice discrimination learning in human *S*s. PSD is the temporal duration of the discriminanda following *S*'s response on both reinforced and nonreinforced trials. The reinforcing stimulus immediately follows correct responses. Perhaps in many learning situations cues are withdrawn with the response

[1] This research was supported in part by Research Grant M-4433 from the National Institute of Mental Health, United States Public Health Service, made to the second author. The first author is a predoctoral NIMH fellow in the Peabody Mental Deficiency program.

and are contiguous neither with the reinforcement nor the terminal phase of the response. Thus, such a task somewhat parallels the trace classical conditioning model, and the formation of an association may depend upon a "memory trace" of the initial event. Ellis (1963) has adduced evidence suggesting that mentally retarded *S*s have inadequate short-term memory processes. Therefore, such *S*s may be more adversely affected by noncontiguity of events than are normal *S*s.

Experiment I

Method

Apparatus.—The *S*s were tested in a 5 × 8 ft. cubicle. The *S* console was a metal cabinet, 16 × 19 in. and 12 in. high, with a top that sloped toward *S*. Two miniature stimulus projectors were mounted 2 in. apart in the sloping top. The viewing surfaces (frosted Plexiglas) of the projectors are round and approximately 1 in. in diameter. Any one of seven patterns or five colors or a combination of pattern and color may be displayed with each projector. A push-button type switch was located below each projector. A green pilot-type lamp mounted 6 in. above the projectors in the console served as a reinforcing stimulus. The console was on a table, and *S* sat in a chair in front of the table. Electromechanical circuitry located in an adjoining room provided automatic control of the intertrial interval, the onset of the stimuli, the PSD, the recording of responses, and the occurrence and duration of the reinforcing stimulus.

Subjects.—The *S*s were 80 undergraduate psychology students and 80 mildly retarded students from a state residential facility. The CAs of the normal *S*s ranged from 17 yr. to 26 yr. with a mean of 20.6 yr. The CAs of the retarded ranged from 15 yr. to 20 yr. with a mean of 16.6 yr. Their IQs ranged from 70 to 85 with a mean of 78.1.

Procedure.—The stimuli were a square, triangle, circle, vertical bar, and horizontal bar presented on backgrounds of blue, green, yellow, red, and light gray, making a total of 25 stimuli. Five were selected arbitrarily as correct, and 1 of the 5 correct and 1 of the 20 incorrect stimuli were presented on each trial. All 5 correct stimuli appeared in each 5-trial block, although incorrect stimuli were assigned at random and did not appear consistently with particular correct stimuli. The serial order and the position of correct stimuli varied randomly within trial blocks. The programing equipment provided for a series of 25 unique trials which was repeated without interruption until *S* either reached a criterion of 10 consecutive correct responses or completed 200 trials.

The *S* was seated before the panel and was told that he was to try to press the button under the "right" picture as many times as he could, that when he pressed the button under the "right" picture the light (reinforcing) would come on, and that when he pressed the button under the "wrong" picture the light would not come on. The *S* was free to respond after the two stimuli appeared. The reinforcing stimulus immediately followed a correct response and was of 1.0 sec. duration. The discriminanda persisted following both reinforced and nonreinforced responses. The PSDs were 0.0, 0.5, 1.0, and 2.0 sec., and *S*s from each of the two groups were assigned equally and without bias to these subgroups. The intertrial interval was a constant 10 sec. and was measured from the offset of the discriminanda on one trial to the onset of the discriminanda on the next.

Results

Number of correct responses in 200 trials was the dependent measure. The *S*s reaching criterion were assumed to perform perfectly beyond the criterion. Figure 1 depicts mean percent correct responses for the eight subgroups. Performance of the normal *S*s is superior to that of the retarded *S*s, $F\ (1, 152) = 15.22$, $p < .001$, and the PSD effect is also statistically significant, $F\ (3, 152) = 6.25$, $p < .001$. The latter effect is due to the superiority of the performances of the 0.5-sec. and 2.0-sec. subgroups over those of the 0.0-sec. and 1.0-sec. subgroups. No other differences reach statistical significance. The interaction between ability level and PSD is not significant. Group curves indicate that the superior performance of the 0.5-sec. and 2.0-sec. subgroups appears within the first 40 training trials and persists

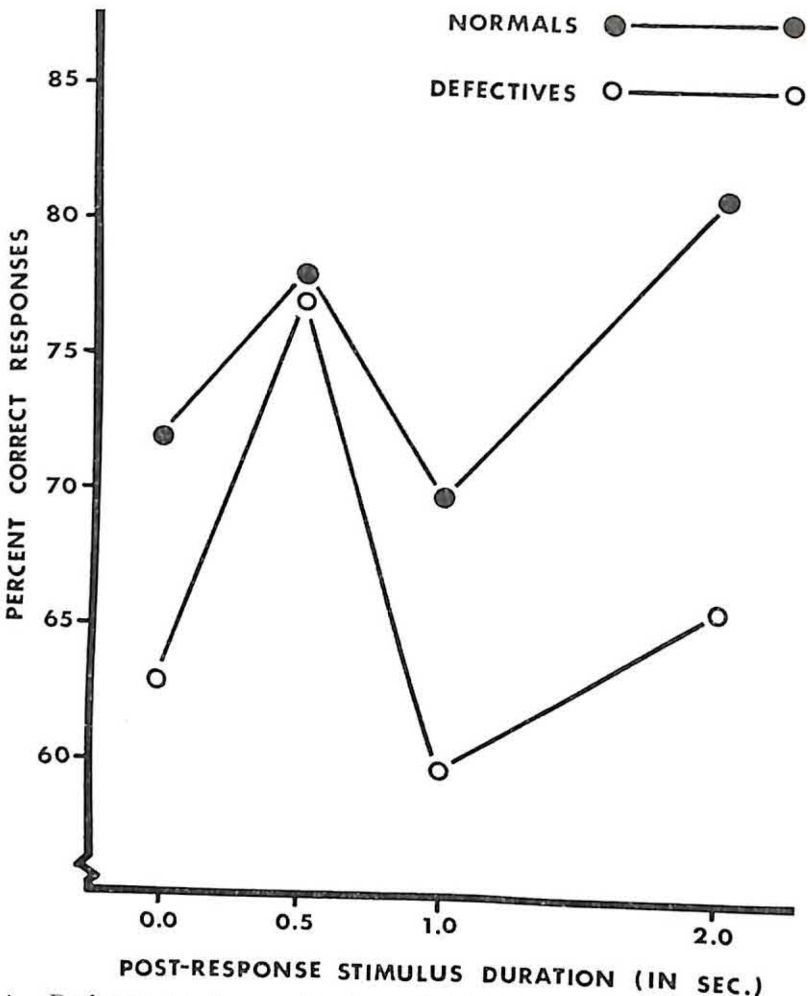


FIG. 1. Performance of normal and retarded *S*s at each of the PSDs in Exp. I.

throughout training. All subgroups show improvement over trials.

These results are not in agreement with general expectations; performance did not increase consistently with PSD, and there is an absence of an interaction effect. Perhaps the most notable effect is the elevated performance in the 0.5-sec. subgroups. The poorer performances of the 1.0-sec. subgroups may have resulted from the simultaneous offset of the discriminanda and the reinforcing light at that PSD (i.e., perhaps some interference occurred).

EXPERIMENT II

Method

Subjects.—The *S*s were 40 undergraduate psychology students and 40 retarded students. None had participated in Exp. I. The retarded students were from a state residential school. The CAs of the normal *S*s ranged from 18 yr. to 27 yr. with a mean of 19.4 yr. The CAs for the retarded *S*s ranged from 15 yr. to 20 yr. with a mean of 16.2 yr., and their IQs ranged from 55 to 85 with a mean of 67.2.

Procedure.—The *S*s from each group were assigned to PSD subgroups of 10 *S*s each. The PSDs were the same as those in Exp. I. The duration of the reinforcing stimulus was extended to 3 sec., and the intertrial interval was measured from *S*'s response on one trial to the onset of the discriminanda on the next. Otherwise, the procedure was identical to that of Exp. I.

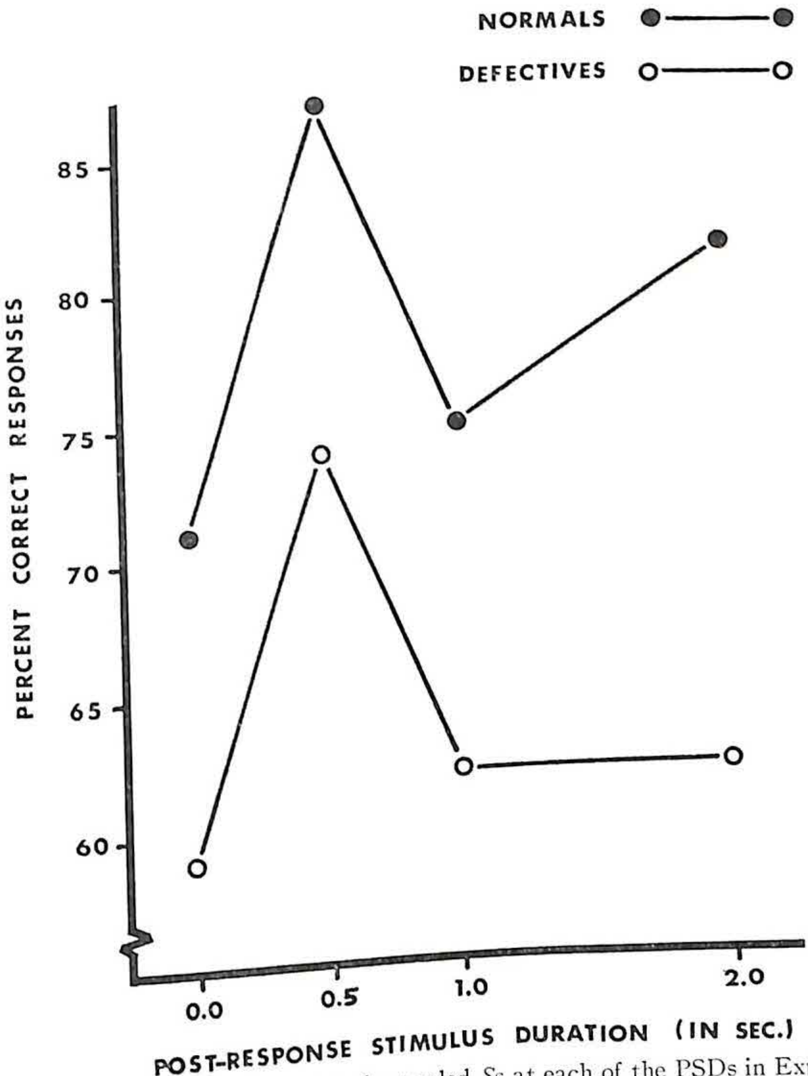


FIG. 2. Performance of normal and retarded *Ss* at each of the PSDs in Exp. II.

Results

Figure 2 shows mean percent correct responses in 200 trials for each of the subgroups. These results are similar to those of Exp. I. Performance of the normal *Ss* is superior to that of the retarded *Ss*, $F(1, 72) = 18.90, p < .001$, and the PSD effect is significant, $F(3, 72) = 4.27, p < .01$. The ability level and PSD interaction is not statistically significant. Group differences again appear early in training and persist throughout.

The general form of the curves in Exp. I and II is quite similar. The effect of the reinforcing stimulus offset is difficult to evaluate since there are some overall group differences between Exp. I and II. However, it apparently produced little, if any, effect. Again, the most marked result is the superior performance of the 0.5-sec. subgroups.

EXPERIMENT III

Method

Subjects.—The *Ss* were 60 undergraduate psychology students. Their CAs ranged from 17 yr. to 25 yr. with a mean of 19.2 yr. None of these *Ss* had participated in Exp. I and II.

Procedure.—Several procedural changes

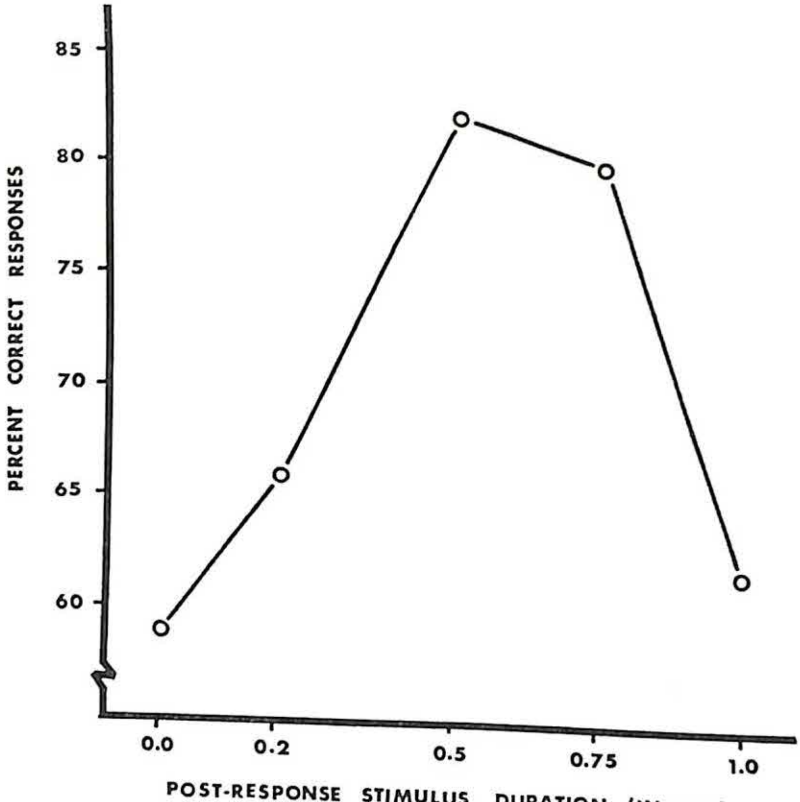


FIG. 3. Performance of *S*s at each of the PSDs in Exp. III.

were introduced to extend the generality of the results of Exp. I and II. The stimuli were six patterns with two identical patterns appearing on each trial, one on a red background and the other on green. For three of the six patterns, green background was correct, and for the other three, red was correct. Each pattern appeared in each 6-trial block. Background color was assigned at random to position, and also serial order of pattern was assigned randomly. The programing equipment provided for a series of 52 trials which could be repeated without interruption.

The reinforcing stimulus was a buzzer sounding for 3 sec. The intertrial interval was a constant 8 sec., and was measured as in Exp. II. The PSDs were 0.0, 0.2, 0.5, 0.75, and 1.0 sec. Ten *S*s were assigned without bias to each of the five groups. Each *S* was practiced to a criterion of 11 correct responses in two consecutive 6-trial blocks or for 102 trials.

Results

Figure 3 presents percent correct responses in 102 trials for the five groups. It is clearly apparent that a PSD of approximately 0.5 sec. yields optimum performance within the time range studied. An analysis of variance of the data yields a significant PSD effect, $F\,(4, 55) = 6.02, p < .001$, and t tests show that the 0.5-sec. and 0.75-sec. groups are superior to the others. This difference again appears early in training and persists throughout. No other differences between groups are significant.

Discussion

These results demonstrate fairly clearly that discrimination learning of the type studied is substantially influenced by PSD. It appears that a PSD of 0.5 sec. yields optimum performance and that longer intervals yield poorer performance up to approximately 1.0 sec. after which it may again increase. The main find-

ings of Sheridan et al. are not in conflict with those of the present experiments since these *E*s found no difference in learning between conditions involving PSDs of 0.0 and 2.0 sec.

The finding that a PSD of 0.5 sec. yields optimum performance is reminiscent of a similar temporal effect deriving from studies of the interstimulus interval in classical conditioning. An interstimulus interval of around 0.5 sec. yields optimum conditioning (Kimble, 1961). However, the relationship between the optimum PSD and the optimum interstimulus interval may be fortuitous since in the present experimental situation the interval between the onset of the CS (discriminanda) and the onset of the UCS (reinforcing light) is determined by the latency of *S*'s response.

Perhaps a number of hypotheses could be summoned to account for the results. None seem particularly parsimonious nor adequate. However, a notion recently expressed by Miller (1963) may prove of value in considering the effects of relatively minute temporal variations in learning tasks. It is conceivable that the "perception" of a stimulus involves a neural change which requires a certain amount of time. The occurrence of another stimulus event while the first is being "processed" may interfere. Such speculation would suggest a more careful analysis of such phenomena as delay of reinforcement. Perhaps *immediate* reinforcement is detrimental to the learning process in comparison with slightly delayed reinforcement.

REFERENCES

CHOW, K. L., & ORBACH, J. Performance of visual discrimination presented tachistoscopically in monkeys with temporal neocortical ablations. *J. comp. physiol. Psychol.*, 1957, **50**, 636–640.

ELLIS, N. R. (Ed.) The stimulus trace and behavioral inadequacy. In, *Handbook of mental deficiency.* New York: McGraw-Hill, 1963.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

MILLER, N. E. Some reflections on the law of effect produce a new alternative to drive reduction. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1963.* Lincoln, Nebraska: Univer. Nebraska Press, 1963.

RIOPELLE, A. J., & ADDISON, R. G. Temporal factors in pattern discrimination by monkeys. *J. comp. physiol. Psychol.*, 1962, **55**, 926–930.

SHERIDAN, C. L., HOREL, J. A., & MEYER, D. R. Effects of response-induced stimulus change on primate discrimination learning. *J. comp. physiol. Psychol.*, 1962, **55**, 511–514.

(Received March 18, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 534–536

SUPPLEMENTARY REPORTS

STIMULUS DURATIONS AND TOTAL LEARNING TIME IN PAIRED-ASSOCIATES LEARNING[1]

CALVIN F. NODINE

Miami University

The present report supplements an earlier investigation concerned with the role of exposure durations of stimulus and response members on the acquisition of a paired-associates list using the anticipation method. Previously, Ss were given a fixed number of learning trials, and exposure durations of stimulus members alone (St duration) and exposure durations of stimulus and response members paired (St-R duration), were found directly related to rate of learning. In the present experiment, Ss' performance was equated by employing a performance criterion, and effects of independent variations in St, St-R durations were assessed using total learning time as the response measure. Total learning time refers to Trials × Exposure Time. The values of both St and St-R durations were $\frac{1}{2}$, 1, 2, and 4 sec. Stimulus durations were found directly related to total learning time, but had a nonsignificant effect on learning rate when Ss were equated in terms of performance. Conversely, St-R durations were found invariant with respect to total learning time, but directly related to learning rate.

In an earlier paper, Nodine (1963) found an eightfold increase in exposure duration of stimulus members alone (St duration) and stimulus and response members paired (St-R duration) from $\frac{1}{2}:\frac{1}{2}$ to 4:4 sec. produced slightly less than an eightfold increase in numbers of correct responses. Total learning time for *S*s in the former exposure-duration combination at the end of 21 trials, however, was 5.5 times greater than that for *S*s in the latter combination. Unfortunately, data limitations precluded further extrapolations of the data, but suggested that some intermediate St, St-R combination might provide an optimum relationship between learning performance and total learning time.

The present experiment extends Nodine's previous findings by assessing the effects of independent variations in exposure duration of stimulus and response members on Total Learning Time (TLT) defined as Trials to Criterion × St, St-R Exposure Time. Of interest also are the effects of these variables on stimulus term (R-S) recall.

Method.—The conditions, apparatus, materials, and procedure of the present experiment differed from those of Nodine's in three details: use of 5 rather than 10 *S*s in each of the 16 St, St-R combinations; *S*s taken to a 15/16 learning criterion rather than a fixed number of trials; and the addition of an R-S recall trial using a 4:4-sec. St, St-R combination immediately following the criterion trial during which only response members were presented and *S*s were required to supply the stimulus member for each pair. The values of St, St-R durations remained the same; $\frac{1}{2}$, 1, 2, and 4 sec.

Results and discussion.—Analysis of variance of correct responses for the first 20 trials in the present experiment (equivalent to Nodine's, 1963, study) revealed a pattern of significant effects similar to those of Nodine's earlier experiment. The effect of St-R durations was highly significant, $F(3, 64) = 47.87$, $p < .01$. Holding St durations constant, a pattern of increasing numbers of correct responses was obtained as St-R durations increased from $\frac{1}{2}$ to 4 sec. This pattern was also present as St durations increased, but less markedly. The effect of St durations, as a consequence, fell short of statistical significance, $F(3, 64) = 1.01$. Correct responses increased significantly from Trial Block 1 to Trial Block 4, $F(3, 192) = 340.45$, $p < .01$. Trials also interacted with exposure durations reflecting increasing increments in numbers of correct responses from Trial Block 1 to Trial Block 4 as both St and St-R durations increased. The slope of the curve over trials for St-R durations was steeper, however, than that for St durations.

[1] This research was supported by Public Health Service Research Grant MH 07828-01, from the National Institutes of Health. The author wishes to express his appreciation to Fred Farzanegan for collecting the data.

Response latencies were found to be directly related to St-R durations. An analysis of variance of reciprocals of latencies revealed a significant effect for St-R durations, $F(3, 64) = 30.70$, $p < .01$ but not for St durations. Latency scores also decreased significantly over trials in a negatively accelerated fashion and interacted with both St and St-R durations.

Figure 1 shows the relationship between TLT and each of the 16 St, St-R combinations. An analysis of variance was performed

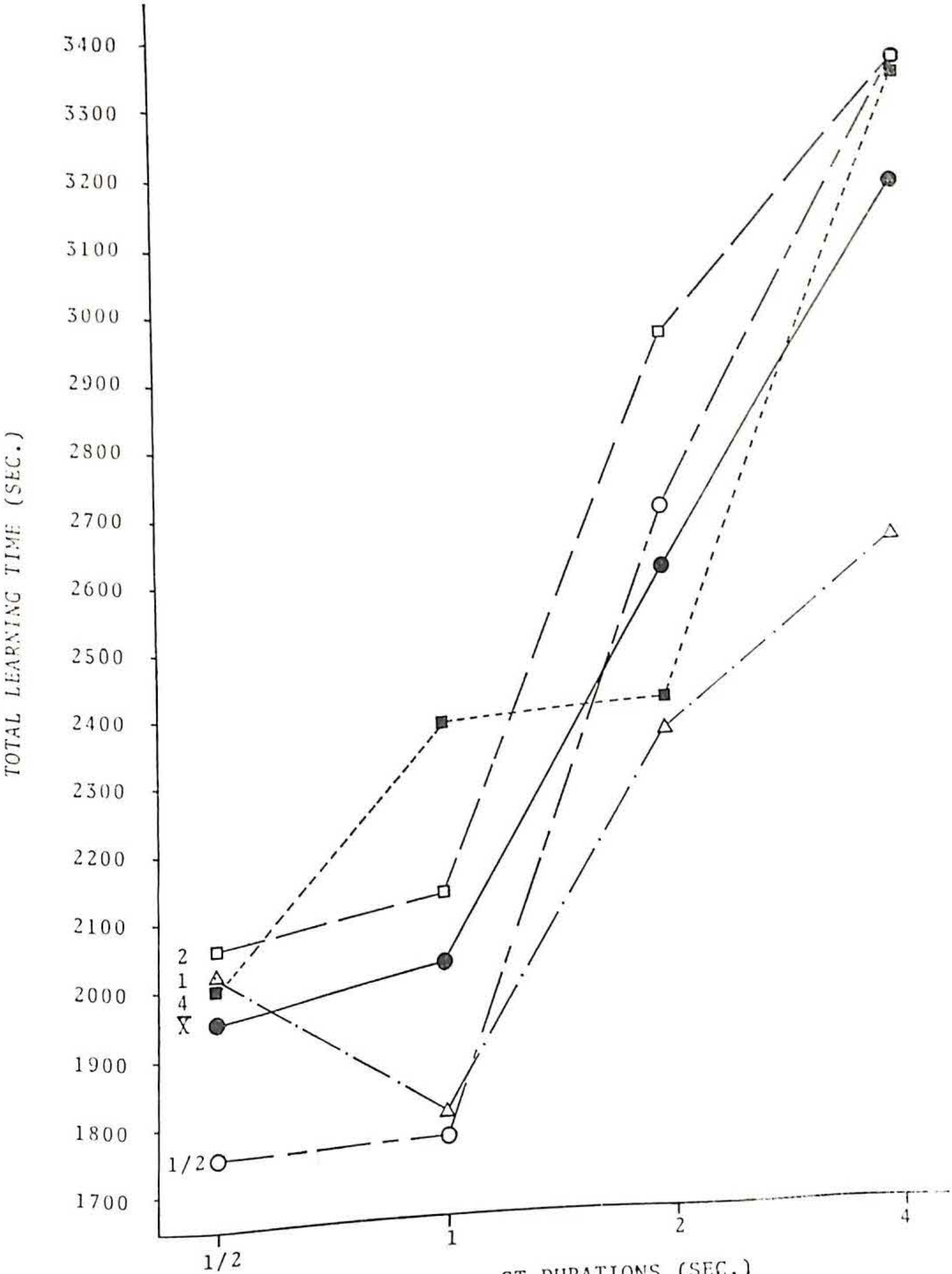


FIG. 1. Mean total learning time for each St, St-R duration. (Separate curves are plotted for each St-R duration. The combined curve for all four St-R durations is indicated by $\bar{X}$.)

on TLT. The F for St durations was highly significant, $F(3, 64) = 15.10$, $p < .01$. Neither St-R nor the interaction was significant. The highly significant F for St durations suggests this factor was the primary contributor to TLT.

A comparison of means for each St, St-R combination using the Tukey procedure showed significant differences between ½- and 4-sec. St durations, $F(4, 64) = 8.29$, $p < .01$, and 1- and 4-sec. St durations, $F(4, 64) = 7.89$, $p < .01$, indicating that ½- and 1-sec. St durations made considerably less contribution to TLT than did either 2- or 4-sec. St durations in combination with St-R durations. A trend analysis of the St duration main effect revealed a highly significant linear component, $F(1, 64) = 41.51$, $p < .01$, suggesting that an increasing linear function best described the relationship between St durations and TLT.

In general, it appears that the ½- and 1-sec. St durations combined with ½- and 1-sec. St-R durations produced the most efficient St, St-R combinations with respect to TLT. From Fig. 1, it is evident that the 1-sec. St-R curve was least affected by St durations. When performance was equated for all combinations in the present study, the ½:½-sec. combination had the shortest TLT. In contrast, when trials rather than performance was equated, Nodine previously found the ½:½-sec. combination ranked last in number of correct responses while the 4:4-sec. combination ranked first. Clearly, from the present findings, the 2:2-sec. combination which is conventionally employed in PA tasks is not as efficient as either the ½-sec. or 1-sec. St, St-R combinations when performance is considered in relation to TLT. It is also evident from these data that St durations are not invariant with respect to TLT. For St-R durations, TLT appears to follow no consistent trend. From these data, it appears that St-R durations are important determiners of learning rate while St durations beyond 1 sec., inflate TLT without accompanying increases in performance. Finally, the analysis of the R-S recalls showed no significant effects suggesting that neither St nor St-R durations influenced the recall of stimulus terms.

REFERENCE

NODINE, C. F. Stimulus durations and stimulus characteristics in paired-associates learning. *J. exp. Psychol.*, 1963, **66**, 100–106.

(Received February 20, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 536–538

COMPOUND STIMULI, DRIVE STRENGTH, AND PRIMARY STIMULUS GENERALIZATION [1]

ALBERT F. HEALEY [2]

University of Massachusetts

Rats were trained to run to food in a straight runway under 23 hr. food deprivation, with a 400-cps tone sounding and a 79-sq. cm. white square in front of the food cup. Generalization test trials were carried out at 2 levels of deprivation (23 hr. and 12 hr.), 2 levels of the tone dimension (400 cps and 200 cps, and 2 levels of the size dimension (79 sq. cm. and 32 sq. cm.). Variation along the 2 stimulus dimensions decreased response strength and altering both dimensions produced more decrement than varying either singly. The latter, compound, effect was apparently of a simple additive sort. Decreasing deprivation time reliably lowered generalization gradients, again in an apparently additive fashion.

Compound stimulus generalization was studied here in the straight runway using visual size and frequency of a pure tone as two dimensions of the CS. Fink and Patton (1953), with rats, and White (1958), with children, found an increasing generalization decrement in operant responding as more CS dimensions were altered, thus supporting Hull's (1952) theoretical expectation. Butter

[1] This research is part of a master's thesis submitted to the Graduate School of the University of Massachusetts, and was completed during the tenure of a National Institutes of Health Predoctoral fellowship. The author is indebted to Seymour Epstein for his support and guidance. Thanks are also due Frank A. Logan for a critical reading of the manuscript.

[2] Now at Princeton University.

(1963), with pigeons, demonstrated a multiplicative compound effect.

Also studied here was the effect of varying drive strength (hours of food deprivation) on generalization along both stimulus dimensions. Brown (1942) and Newman (1955), both with rats in a straight runway, reported changes in the heights of gradients with alterations in drive level while Hull's (1952) $\bar{H} \times D$ formulation requires both height and slope change with drive level. Both predicted outcomes were obtained by Porter (1962) in the human eyelid-conditioning situation.

Method.—The apparatus was a straight runway, 24 in. long, 4 in. wide, and 6 in. deep, with a separate 12-in. start box. At the goal end was a metal panel with a 1½-in. square door in its center allowing access to a food cup. The door, and part of the area around it were painted flat white to form a 32-sq. cm. white square; a second panel could be slid in front of the first to form a 79-sq. cm. white square. Alongside the runway was a mounted 4-in. loudspeaker which emitted a pure tone of either 200 cps or 400 cps at approximately 85 db.

One second after initiation of the tone, the start door opened and a timer started. This timer stopped and a second started when *S* interrupted a beam 1 in. from the start door. The second timer and the tone ceased when *S* pushed in the door on the stimulus panel.

The *S*s were 48 white albino rats, 100–110 days old at the outset. They were fed for 1 hr. each day with water always available. After adjustment to the apparatus, *S*s were trained to make the door-pushing response to obtain a small, approximately constant, amount of wet mash. For acquisition trials, the tone was set at 400 cps, the 79-sq. cm. white square was in place and *S*s were 23 hr. deprived. Two trials a day were given for 7 days and 4 a day for the next 10 days to total 54, at which point performance had stabilized. To test for generalization, four matched groups of 12 *S*s each were selected. Two groups were tested under the training deprivation level (23 hr.) and two shifted to 12 hr. deprivation by interpolating a 1-hr. feeding session midway between the last training day and the first test day. One of the two groups at each deprivation level was tested with the same tone as in training (400 cps), the other with a 200-cps tone.

With temporal order counterbalanced within each of the four groups, all *S*s were tested with the training square (79 sq. cm.) and, on a second test day, with the smaller

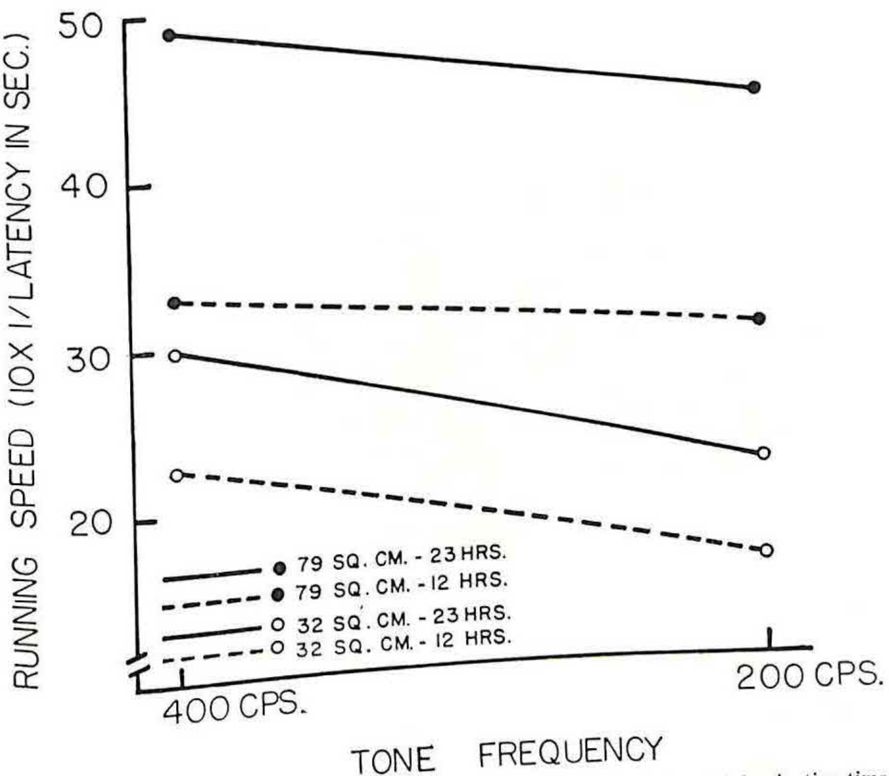


FIG. 1. Mean running speeds under combinations of levels of visual size and deprivation time, as a function of tone frequency.

square (32 sq. cm.). Two recovery days, with acquisition conditions reinstated, intervened between the 2 test days. On each test day *S*s were first given three training trials as usual, then three nonrewarded test trials.

Results and discussion.—Results are reported only for running speed (1/latency) since starting-speed data yielded essentially the same effects. The *F*s included are from an analysis of variance with two between-*S* variables, deprivation level and tone frequency, and repeated measurements with the size dimension and the three test trials. Figure 1 displays mean running speed averaged over test trials, for four combinations of levels of deprivation time and size as a function of tone frequency.

Generalization decrements occurred with variation along the size dimension, F (1, 44) = 20.89, $p < .001$, and the tone dimension, F (1, 44) = 4.55, $p < .05$. Varying both size and tone produced greater response decrement than varying either alone, confirming the theoretical expectation. The contrast for size alone vs. tone plus size, using a formula for independent means, yields a t (46) = 2.35, $p < .05$. With a formula for correlated means, comparison of tone alone vs. tone plus size yields a t (23) = 6.32, $p < .01$. The stimulus dimensions did not interact, a finding at variance with Butter's (1963) results and in agreement with the White (1958) study.

Consistent with studies cited above, decreasing deprivation time reliably decreased test-trial performance, F (1, 44) = 9.89, $p < .005$. Also, with Brown (1942) and Newman (1955), there was no significant interaction of deprivation level with either the tone or the size dimension, although there was an observable trend in that direction with the size dimension (Fig. 1).

REFERENCES

Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **38**, 209–226.

Butter, C. M. Stimulus generalization along one and two dimensions in pigeons. *J. exp. Psychol.*, 1963, **65**, 339–346.

Fink, J. B., & Patton, R. M. Decrement of a learned drinking response accompanying changes in several stimulus characteristics. *J. comp. physiol. Psychol.*, 1953, **46**, 23–27.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Newman, J. R. *Stimulus generalization as a function of drive strength.* (Doctoral dissertation, University of Illinois) Ann Arbor, Mich.: University Microfilms, 1955, No. 55-1332.

Porter, J. J. Stimulus generalization as a function of UCS intensity in eyelid conditioning. *J. exp. Psychol.*, 1962, **63**, 311–313.

White, S. W. Generalization of an instrumental response with variations in two attributes of the CS. *J. exp. Psychol.*, 1958, **56**, 339–343.

(Received February 24, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 538–540

A FAILURE TO FIND A RESPONSE PERSISTING IN THE APPARENT ABSENCE OF MOTIVATION [1]

MILTON A. TRAPOLD, SANDRA R. BLEHERT [2], AND THOMAS STURM

University of Minnesota

This experiment is a replication of an earlier one by Robinson (1961) in which, subsequent to training rats to make both a locomotor and a bar-holding response to terminate the same shock-paired stimulus, extinction of the locomotor avoidance response was shown to produce no transferred extinction of the bar-holding response. This finding is contrary to prediction from the 2-process theory of avoidance learning. The results of this replication were contrary to Robinson's results in that extinction of the locomotor response here produced a large and persisting decrement in bar holding.

Robinson (1961) has recently reported results which are seriously at variance with the widely accepted two-process theory of discriminated avoidance learning (Kimble, 1961). Briefly, Robinson's experiment involved (*a*) training rats to make two different responses to terminate the same shock-paired CS, (*b*) extinguishing one of these responses in the experimental group but not in the control group, and (*c*) comparing the two groups on their performance over a series of extinction sessions on the second response. He found that the prior extinction of one

[1] This research was supported by a grant from the Graduate Research Fund, University of Minnesota.
[2] Now at Stanford University.

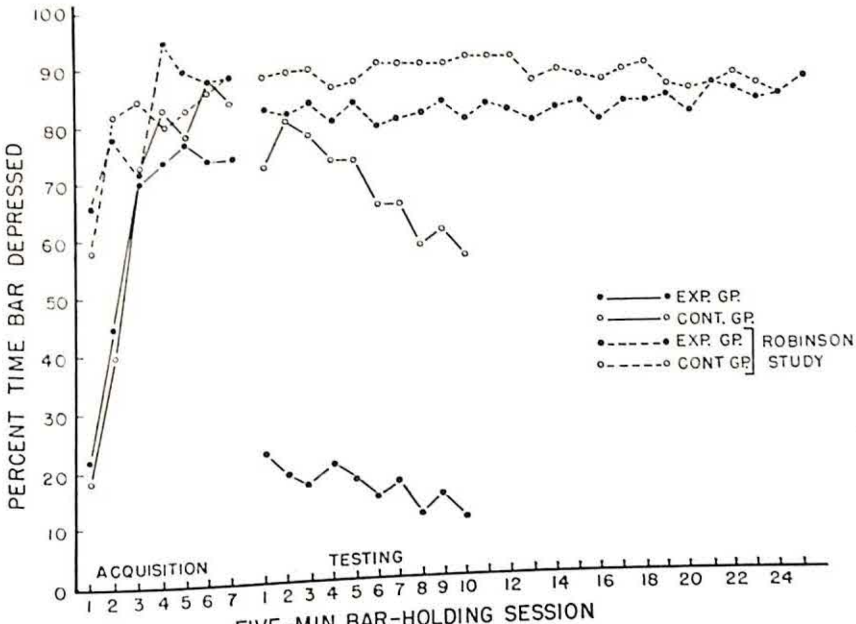


FIG. 1. Mean percentage of time *S*s held the bar down to keep the shock-paired CS off during acquisition and testing for the experimental and control groups of this study and of Robinson's (1961) original study. (The experimental and control groups from the Robinson study were there labeled I_A and I_B, respectively, and the data presented here were estimated from Fig. 2 and 4 of that report.)

response in the experimental group did not, as would be expected from the two-process theory, lead to faster subsequent extinction of the second response. This experiment is an attempt to replicate as closely as possible the critical groups of Robinson's study.

Method.—The *S*s were 15 male and 15 female naive hooded rats 100 days old. The apparatus was an automated shuttle box comprised of two identical compartments 18 in. long, 8 in. wide, 12 in. high, with grid floors for scrambled shock, metal walls, and separated by a permanent metal barrier to 3.5 in. from the floor, and by a guillotine door from the barrier to the ceiling.

The procedure involved first training all *S*s to jump from one compartment to the other during a 5-sec. CS (light + buzzer + open guillotine door) to avoid shock. The details of the training procedure were identical to Robinson's with the exception that here the shock was a constant 375 v. ac (.2 megohms in series with *S*) for all *S*s rather than the "just subtetanizing" shock level, individually determined for each *S*, employed by Robinson. Then, with the doorway between compartments blocked by a transparent barrier, all *S*s were confined to one compartment and trained to hold a bar depressed to keep the CS off. The procedure here was again identical to Robinson's except for the use of small amounts of shaping on *S*s having low operant bar-press rates.

The experimental group ($N = 17$) was then subjected to extinction of the hurdle-jumping response using Robinson's procedure exactly but with a limit of 24 extinction sessions if *S* failed to reach the extinction criterion. The control group ($N = 9$) remained in their home cages during this phase. All *S*s were then given 10 bar-holding test sessions under conditions identical to bar-holding acquisition.

Results and discussion.—The experimental and control groups did not differ in the acquisition of either the hurdle response or the bar-holding response. There were, however, contrary to Robinson's results, gross differences between these two groups on the bar-holding tests after extinction of hurdling in the experimental group. These differences appear in Fig. 1 where the mean percentage of time the bar was depressed in a 5-min. session is plotted over sessions for *S*s of this experiment and for the comparable *S*s of Robinson's study. The slight difference between Robinson's groups was not significant over the 25 tests, nor was there any evidence of extinction in either of his groups. The experimental group of this study, on the other hand, showed a significant decrease from the last 2 days of acquisition to the first 2 days of testing, $t\,(16) = 8.04$, $p < .001$; and was significantly below the control group over all 10 test sessions; $t\,(24) = 6.21$, $p < .001$. Moreover, the control group here showed a

significant amount of extinction when the first three test sessions are compared with the last three; t (8) = 2.94, $p < .05$.

The results of this study are consistent with the two-process theory of avoidance learning, but are clearly inconsistent with Robinson's data. Moreover, no good explanation for the latter is apparent. The only two procedural differences of note were the use here of a constant shock level, rather than individually determined "just subtetanizing" shocks, and the necessity here to use some shaping early in the acquisition of the bar-pressing response (which Robinson does not report having to use). It does not seem likely, however, that either of these differences is responsible for the differences in results.

The change in shock was necessitated when in some preliminary research, several *S*s were damaged because of large fluctuations, both within and between sessions, in their tetany threshold. Thus, when a thorough review of the shock-intensity literature revealed no suggestion of any qualitative differences between shock values in the "just subtetanizing" range and those slightly less intense, a fixed shock value slightly below the lowest ever observed to produce tetany was substituted.

Similarly, the writers are aware of no evidence to suggest that shaping would produce a response so much more susceptible to transferred extinction effects, and this is further borne out by the absence in the present data of any systematic differences between *S*s who were or were not shaped.

There were, additionally, slight differences between this study and Robinson's in the construction of the apparatus, the composition of the CS, and the strain of rats, but again there is no good reason to think that the grossly different results were due to any of these, and no other explanations are apparent to the writers.

REFERENCES

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

ROBINSON, H. B. Persistence of a response in the apparent absence of motivation. *J. exp. Psychol.*, 1961, 61, 480–488.

(Received February 26, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 5, 541–542

Beginning with the first issue of Volume 69 and continuing through the remainder of that Volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are the entire backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication. The articles listed below are scheduled to appear approximately 12 months hence.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Discrimination Shifts as a Function of Degree of Training in Children: James Youniss* and Hans G. Furth: Department of Psychology and Psychiatry, Catholic University of America, Washington 17, D. C.

Stimulus Fluctuation, Reactive Inhibition, and Time between Trials in Classical Eyelid Conditioning: William F. Prokasy*: Department of Psychology, 117 Burrowes Building, Pennsylvania State University, University Park, Pennsylvania 16802.

Dependence of Equality Judgments upon the Temporal Interval between Stimulus Presentations: Wallace R. McAllister,* Dorothy E. McAllister, and Joseph J. Franchina: Psychology Department, Syracuse University, Syracuse, New York 13210.

Task Characteristics in Sequential Decision Behavior: William C. Howell*: Laboratory of Aviation Psychology, Ohio State University, 1314 Kinnear Road, Columbus, Ohio 43212.

Incubation and Inhibition: Sanford Golin* and Anne Keefe Golin: Department of Psychology, University of Wisconsin, 3203 Downer Avenue, Milwaukee, Wisconsin 53211.

Visual and Pronouncing Responses, and the Relation between Orienting Task and Presentations in Incidental Learning: Arnold Mechanic* and Joanne D'Andrea: Division of Science and Mathematics, California State College, Hayward, California 94542.

Vigilance Performance with a Qualitative Shift in Verbal Reinforcers: William Bevan* and Edward D. Turner: Anderson Hall, Kansas State University, Manhattan, Kansas 66504.

Effect of Amount of Pretraining with Identical and Dissimilar Stimuli on Concept Learning: Richard D. Petre*: Bureau of Child Research, Children's Rehabilitation Unit, University of Kansas Medical Center, 39th and Rainbow Boulevard, Kansas City, Kansas 66103.

Resistance to Extinction as a Function of Percentage of Reinforcement, Number of Training Trials, and Conditioned Reinforcement: Norman Kass* and Helen Wilson: Department of Psychology, San Diego State College, San Diego, California 90015.

Persistence to Continuous Punishment Following Intermittent Punishment Training: R. K. Banks*: Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

Some Determinants of the Use of Relationships in Discrimination Learning: Peder Johnson and Daniel E. Bailey*: Department of Psychology, University of Colorado, Boulder, Colorado 80304.

Resolving of Successive Clicks by the Ears and Skin: George A. Gescheider*: Department of Psychology, Hamilton College, Clinton, New York 13323.

Stimulus Generalization of a Positive Conditioned Reinforcer: III. The New Learning Method: Salvatore C. Caronite and David R. Thomas*: Department of Psychology, Kent State University, Kent, Ohio 44240.

Diminution and Recovery of the UCR in Delayed and Trace Classical GSR Conditioning: Ronald Baxter*: Pittsburgh Child Guidance Center, 201 De Soto Street, Pittsburgh, Pennsylvania 15213.

Context Stimuli in Verbal Learning and the Persistence of Associative Factors: Isabel M. Birnbaum*: Institute of Human Learning, 2241 College Avenue, Berkeley, California 94720.

---

* Asterisk indicates author for whom address is supplied.

Effects of Differential Instructions, Differential Payoffs, and Presence or Absence of Feedback on the Percentage, Latency, and Amplitude of the Conditioned Eyelid Response: Harold D. Fishbein* and I. Gormezano: Department of Psychology, University of Cincinnati, Cincinnati 21, Ohio.

Optimality of Perceptual Decision Criteria: Z. Joseph Ulehla*: Department of Psychology, University of Denver, University Park, Denver, Colorado 80210.

Transfer of Discrimination: Frank A. Logan*: Department of Psychology, University of New Mexico, Albuquerque, New Mexico.

Extinction of the Human Eyelid CR as a Function of Presence or Absence of the UCS during Extinction: Kenneth W. Spence*: Department of Psychology, Mezes Hall 211, University of Texas, Austin, Texas 78712.

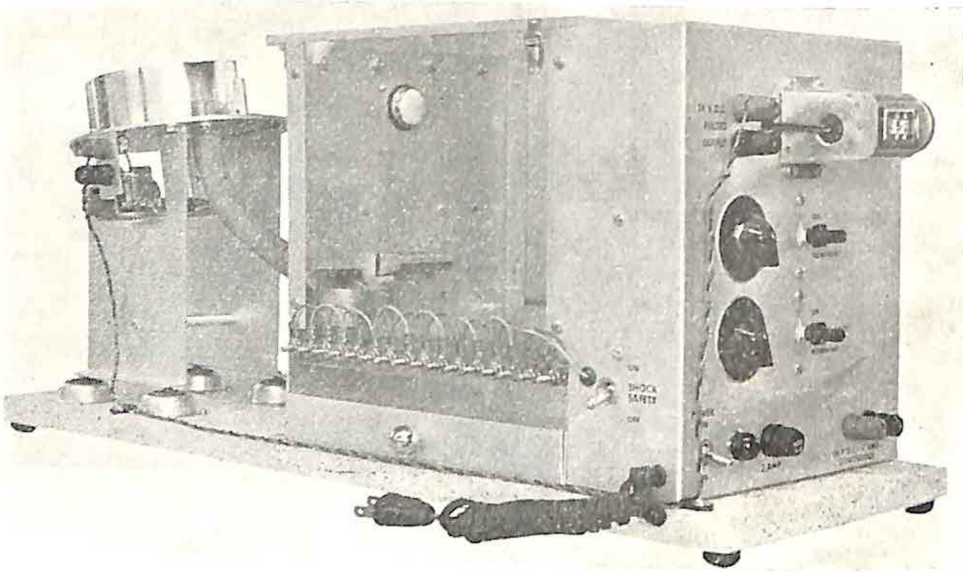
Effect of Perceptual Pretraining on Reversal and Nonreversal Shifts: Louise Sherlock Tighe*: Department of Psychology, Dartmouth College, Hanover, New Hampshire.

Individual and Group Predictions of Free Learning: Benton J. Underwood*: Department of Psychology, Northwestern University, Evanston, Illinois.

Generalization of Serial Position in Rote Serial Learning: R. Yale Hicks, David T. Hakes,* and Robert K. Young: Department of Psychology, Mezes Hall 211, University of Texas, Austin, Texas 78712.

Effect of Instructional Set and Amount of First Learning on Negative Transfer: Wilma A. Winnick*: Department of Psychology, Queens College of the City University of New York, Flushing, New York 11367.

NEW MODEL A-105 ANIMAL CAGE



Student Lab or Demonstration Box Animal Cage with Many Features by Scientific Prototype

This new Scientific Prototype animal cage combines cage, pellet feeder, power supply, light and shock source intensity controls, and counter. The unit is completely wired, mounted on a base, and ready to operate. The rat box is approximately 9 × 8⅝ × 8. Grids are stainless steel. Top and side are ¼″ clear Plexiglass. The stainless steel pedal lever is counterbalanced to permit sensitivity adjustments. Dropping tray is aluminum anodized. Of special interest is the feature of self-contained cue light and shock source *intensity controls*. Price . . . $259.50 F.O.B. N.Y.C. . . . quantity discounts. Full information available.

NEW LDR SWITCH. Provides transient free audio switching with the lowest distortion ever offered.

NEW DIGITAL EQUIPMENT CATALOG

Covers theory and operation of digital equipment for laboratory or clinical work. Shows and discusses Scientific Prototype's complete line of power supplies, gates, multivibrators, clocks and counters, output units, and special purpose units.

SCIENTIFIC PROTOTYPE MFG., CORP.

Behavior Analysis Apparatus . . . operant conditioning kits, rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

615 West 131 St., New York, New York
Telephone: (212) 368-6000

Journal of Experimental Psychology

Vol. 69, No. 6 — June 1965

PARAMETERS OF PAIRED-ASSOCIATE VERBAL LEARNING: LENGTH OF LIST, MEANINGFULNESS, RATE OF PRESENTATION, AND ABILITY[1]

JOHN B. CARROLL AND MARY LONG BURKE

Harvard University

144 10th-grade Ss learned verbal paired associates in a factorial design with 4 list lengths (4, 8, 12, and 20 pairs), 3 levels of meaningfulness, 3 rates of presentation (3, 4, and 8 sec.), 2 levels of ability, and 2 equivalent forms. Dependent variable was amount of time to learn each pair to 1 correct anticipation, summed over the pairs in a list; logarithms of these measures were used. All main effects except list form were significant. For the 9 combinations of Meaningfulness × Rate of Presentation, the regressions of $\log_e$ learning times on $\log_e$ list length did not depart significantly from linearity; they had a common slope but significantly different intercepts, in agreement with Thurstone's rational equation stating that learning time increases as the 3/2 power of list length.

The literature of paired-associate learning yields little information concerning optimal rates of introducing new items when there are large numbers of items to be learned. The present study was designed to investigate some of the important parameters of paired-associate learning in the classical anticipation-method experiment. It was hoped that the results might suggest what sorts of schedules and rates of introducing new items might be usefully experimented with in the context of programed instruction.

Search of the literature revealed that little investigation of list length as a parameter in verbal learning had been done since Thurstone's (1930) formulation of a rational equation relating learning time and list length. It was felt desirable to extend the work of Cieutat, Stockwell, and Noble (1958) on meaningfulness and learning ability and of Hovland (1949) on rate of presentation by incorporating list length, meaningfulness, rate of pres-

[1] This research was supported in part by the Committee on Programed Instruction, Harvard University, under a grant from the Carnegie Corporation of New York. Thanks are due to Donald Davidson, Principal of Newton South High School, for making students from his school available to serve as experimental Ss. We are grateful to Albert E. Beaton, Jr. for help with certain aspects of the computations.

entation, and ability level in a single factorial experiment. At the same time, an effort was made to avoid complexity by selecting stimulus and response terms at the same levels of meaningfulness, setting presentation (S term only) and confirmation (S term plus R term) periods equal, and eliminating all intertrial rest periods. The problem posed was to attempt to formulate a general equation that would permit the estimation, as accurately as possible, of learning time as a function of the four variables selected for study.

Method

Subjects.—The *S*s were 144 students in a large suburban high school, randomly selected from the entire tenth-grade class. Boys and girls were about equally represented. In study hall periods, *S*s were given Part V, Paired Associates, of the Modern Language Aptitude Test (MLAT) (Carroll & Sapon, 1958) and divided into two groups—fast and slow learners—at the group raw score median (10.5). Kjeldergaard (1962a) had demonstrated that use of this test could increase the efficiency of experiments involving paired-associate learning.

Materials.—The practice list (see below) consisted of 10 pairs of trigrams with association values of 70–77% (Archer, 1960): WEY-LAV; MUL-KEP; DAF-PIR; VOT-YAW; SIV-HUS; FAS-TEK; CEN-BIP; PAG-DUX; RUD-VIK; NAC-FEN.

The lists of paired associates used in the main experiment were at three levels of meaningfulness: High, Medium, and Low. Pairs for the high-meaningfulness lists were pairs that had been missed by fewer than 50% of a group of high-school and college students in a "one-trial learning" experiment conducted by Kjeldergaard (1962b); the words in these pairs were nouns, verbs, or adjectives of varying length with frequency ratings of AA or A in Thorndike and Lorge (1944). Medium-meaningfulness lists used low-frequency five-letter words that occurred at least .5 times, but less than 2 times per million on the Lorge magazine and semantic counts and on the Thorndike general count. Low-meaningfulness materials were trigrams with low (0–35%) association values (Archer, 1960).

In the construction of lists, stimulus and response words were paired randomly under the restriction that members of a pair should not begin with the same letter or terminate with the same combination of letters. Furthermore, in any given list, it was true of both the stimulus terms and the response terms that no two terms began with the same letter or ended in the same combination of letters (with one or two exceptions, e.g., EIDER and UMBER).

Within each level of meaningfulness, lists of 4, 8, 12, and 20 pairs were used. Twelve-pair lists were constructed by random selection of pairs from the 20-pair lists, 8-pair lists from the 12-pair lists, and 4-pair lists from the 8-pair lists. There were two forms of each list, containing entirely different pairs (Forms A and B) to minimize the possibility of list-specific results. The lists are shown in Table 1.

For use in the Lafayette memory drum (Model 302B), each list was prepared in the form of three different random orders printed in capital letters (from an IBM 402 tabulator) on a continuous loop of paper. Presentation of each random order was considered as constituting one trial; trials succeeded one another without any break.

Procedure.—The *S*s were brought to the experimental room individually and instructed in the use of the memory drum. The anticipation method was used throughout the experiment; *S* was to try to say aloud (pronounce) the correct response before it was exposed by the shutter in the right-hand section of the window in the memory drum. The *S*s were allowed to use any reasonable pronunciations of the response terms. For warm-up, each *S* was given 15 trials on the practice list at a 4-sec. presentation rate (2 sec. for presentation of the stimulus term alone, and 2 sec. for presentation of the response term along with the stimulus term). Each *S* was then required to learn a particular experimental list at a particular rate of presentation—3, 4, or 8 sec. per pair, the time being equally divided between the stimulus presentation and the stimulus-response presentation. The criterion of learning was one perfect trial, but the experiment was terminated after 40 trials if this criterion had not been achieved. The correctness of each anticipation was noted, a failure to anticipate being recorded as incorrect.

Design.—The materials and procedures described above implied a $4 \times 3 \times 3 \times 2 \times 2$ factorial design; each of the 144 *S*s was tested under a unique condition. Preliminary analyses of the results indicated that there

TABLE 1

EXPERIMENTAL LISTS

No. of Pairs	High Meaningfulness		Medium Meaningfulness		Low Meaningfulness	
	Form A	Form B	Form A	Form B	Form A	Form B
4	PAIN COLD WALK ISLAND OPERATION GENERAL HUSBAND DRAW	HAND FULL ISSUE NINE OLD JOHN NURSE PAY	DELFT MERLE FAGOT WOOER EGRET KETCH UNCUT ANODE	WHELK ILIUM FACET LEACH BASIL OVATE SWALE FIBRE	FEP RUX ZEM GIX YUN FEG NAF ZOD	FYM RUV ZER GEX YIT FIK NAX ZAB
8	NEWSPAPER THOUSAND GAVE MET SEAT DOCTOR NATION GO	VAIN GLAD ESCAPE WIN CAPITAL GUARD PIECE OFFER	SOUSE LLANO PHIAL NADIR YUCCA TRYST TAUPE REFIT	LIMBO VISOR INGOT CAPON TUNNY MANSE NONCE ETUDE	VEZ LYG BEX JUF HUV KIF TEY YAG	TAH YEM HAJ KIV BOJ JUB VOP LEQ
12	TWELVE LEG VILLAGE GATHER EARTH INCREASE TRAVEL CLOUD	CORNER EAR ACCEPT TURN RECORD LIP SECOND GRAIN	GRUEL EIDER LEAPT UMBER WINCH HOURI AXIOM OATEN	PAYER ALACK REBEC GENIE EERIE NATAL VENAL BLEAR	GAC TUZ RUH VAW SIJ DYB KEB SEJ	KIG SAJ RYX VUS GOX TAZ SIW DEJ
20	MOUTH HAIR CREATE STRANGE WARM HOLD KNOWN INCH PAST LEFT NIGHT MANNER STAND MEAN FIGHT LAKE	REACH JOIN DETAIL EXPECT PLANT TOUCH OCEAN SHOT SIT PUBLIC FASHION GENTLE DREAM HIGH LAW DESTROY	CAIRN GNOME MIDGE DAVIT JABOT IDIOM RANGY SAHIB ORATE BEDEW BIGHT PRAWN KULAK FLUME NITRE JUNTO	GROAT SULLY CLACK JOIST ODIUM KRAAL DUCAT RENAL ARRAS TRIPE KRONE HORNY MILCH DOUSE JUNTA PRONG	LAJ MUP JOZ NAQ QAZ COG PEF BAV CUG WIB MIV QEN DUJ POB WOX HYF	PYB BOF JID NUK DYF PIB QOT CEG WYD HIJ CEF WUG LUJ MEB MYP QUM

Note.—Each list of length $L=4$, 8, 12, or 20, is presented here as the first L pairs in a column. The response term in this experiment was always the second member of the pair.

were no significant differences ($p > .05$) between the two forms of any list with respect to number of trials to criterion (40 being used as the score for those *S*s not attaining a perfect trial), and thus list form was employed as a within-cell variable in subsequent analyses.

Results

Number of trials to criterion.—In all, 32 out of the 144 *S*s failed to achieve criterion in 40 trials. It is evident from Table 2 that low meaningfulness and list length were the chief factors associated with failure to achieve criterion. A four-way analysis of variance using as the dependent variable the number of trials performed by each *S* (whether or not he attained criterion) showed all main effects significant ($p < .01$), with no significant interactions. Because of the ambiguity of this dependent variable, however, these results deserve little attention.

Time scores with missing data supplied by extrapolation.—Time to learn a specified number of pairs was con-

TABLE 2

PERCENTAGE OF *S*s FAILING TO REACH CRITERION IN 40 TRIALS, BY LEVELS OF MAIN EFFECTS

Meaningfulness[a]	%	List Length[b]	%	Presentation Rate[c] (Sec.)	%	Ability[d]	%
High	6.2	4	2.8	8	18.7	Fast	20.8
Medium	14.6	8	16.6	4	14.6	Slow	23.6
Low	45.7	12	27.8	3	33.3		
		20	41.6				

[a] $N = 48$ each level.
[b] $N = 36$ each level.
[c] $N = 48$ each level.
[d] $N = 72$ each level.

sidered a more meaningful and useful measure of learning than number of trials to criterion. In the case of *S*s who failed to attain the overall criterion in 40 trials, it was necessary to estimate by extrapolation the number of trials that they would have required to attain a given number of pairs (i.e., to have correctly anticipated each of that number of pairs at least once).[2] It was observed that for any given *S* there was an approximately linear relation between the number of items learned (i.e., correctly anticipated at least once) by a given trial and the logarithm of the number of that trial: if this relation were perfect it would be described by the equation

$$\log t = q + bn_t \qquad [1]$$

where t = trial number, n_t = the number of items anticipated correctly at least once at any time from Trial 2 through Trial t, and q and b are parameters pertaining to the individual *S* and the particular conditions of the experiment. For each *S* who had not correctly anticipated every item at least once by the time he terminated the experiment, a least-squares fit (minimizing the sum of the squares of the differences between predicted and actual trial numbers expressed logarithmically) to the above equation was made for all available data points (t, n_t) and the missing data points supplied by extrapolation. Extrapolation was found to be necessary in only 21 cases, all having list lengths of 8, 12, or 20: in only 8 cases were more than three data points missing. Over the 144 cases, the median r between n_t and $\log t$ was .95.

From the data points (t, n_t)—whether actual or extrapolated—there was further computed the total amount of time, T, required by each *S* to learn the first 4 pairs he learned, and where appropriate, the first 8, 12, and 20 pairs learned. (For example, in the case of *S*s who learned 12-pair lists, the amounts of time to learn were computed for the first 4 and the first 8 pairs they happened to learn as well as the amount of time to learn all 12 pairs in the list.) This computation, patterned after a procedure used by Bugelski (1962), totals the amounts of time during which each of a given set of pairs has been presented up to and including the trial on which the pair is correctly anticipated for the first time. For example, an individual who required 5, 9, 10, and 16 trials, respectively, to make correct responses for the first four items he acquired in an 8-pair list presented at the rate of one pair every 3 sec. would be reckoned as having required $3(5 + 9 + 10 + 16) = 120$ sec. for this learning. Logarithms (to the base e) of these time scores were taken in order to secure arrays with more symmetrical distributions than would have been the case with the original time scores in seconds. The resulting logarithmic scores were used for a series of analyses of variance, to be described below.

Of primary interest are the means displayed in Table 3, for these report how much time is required for learning (to a criterion of one correct anticipation each) a given number of pairs in a given experimental setting. Each value entered in the table is a mean for four cases, the data for the two list forms, A and B, and for the two learning ability levels being pooled. It is evident that the amount of time

[2] It should be noted that this is a slightly different criterion from that of a perfect trial for that number of pairs. Several *S*s who did not meet the criterion had even before the fortieth trial correctly anticipated every item at least once during the course of the experiment.

TABLE 3

MEAN LOG$_e$ TIMES (SEC.) TAKEN TO ACQUIRE TO A CRITERION OF ONE CORRECT RESPONSE, THE FIRST 4, 8, 12, OR 20 PAIRS LEARNED, FOR LISTS OF SEVERAL LENGTHS AND LEVELS OF MEANINGFULNESS PRESENTED AT THREE RATES

List Length	Meaningfulness								
	High			Medium			Low		
	Rate of Presentation (Sec.)								
	3	4	8	3	4	8	3	4	8
First 4 Pairs									
4	3.37	3.71	4.19	4.78	4.90	4.96	4.91	4.65	5.22
8	3.51	3.55	4.16	4.67	3.75	4.73	4.68	4.81	5.15
12	3.34	3.57	4.16	4.43	4.28	4.74	4.54	4.54	5.28
20	3.36	3.72	4.21	4.19	4.12	4.86	4.64	4.12	4.85
First 8 Pairs									
8	4.50	4.60	5.05	5.99	5.03	5.75	6.12	5.93	6.13
12	4.37	4.44	5.01	5.43	5.28	5.81	5.77	5.62	6.30
20	4.19	4.59	5.09	5.12	5.06	5.80	5.66	5.38	5.95
First 12 Pairs									
12	5.10	5.05	5.69	6.11	6.11	6.47	6.55	6.44	7.15
20	4.83	5.15	5.65	5.75	5.62	6.41	6.17	5.70	6.68
20 Pairs									
20	5.77	6.11	6.52	6.79	6.48	7.38	7.47	6.74	7.80

Note.—Each entry is the mean of four values (from two alternate forms of each list taken by *S*s at two levels of ability).

required to learn a certain number of pairs is relatively independent of the length of the list in which these pairs are being learned. The mean log$_e$ times required to learn, say, 4 pairs of high-meaningfulness syllables at a 4-sec. presentation rate lie within the range 3.55–3.72 whether the syllables are in a list by themselves or imbedded in a list of 20 syllables; the mean log$_e$ times to learn 8 pairs under similar conditions fall between 4.44 and 4.60; for 12 pairs, between 5.05 and 5.15; and for 20 pairs, the mean log$_e$ time is 6.11. This would suggest that there is a uniformity in the rate at which new information can be acquired that largely transcends the context in which the learning takes place.

Data presented in Table 3 suggest, however, that the mean log$_e$ time to learn the first n pairs tends to decrease somewhat with the length of the list in which these pairs are imbedded. In four-way ANOVAs performed for each of the first three blocks of Table 3 (ability level being also used as one of the main effect variables), list length was a significant effect at either the 1% (for 4 pairs and for 12 pairs) or the 5% for 8 pairs) levels. This is probably due to the sort of selection artifact noted by Williams (1961): on the assumption that an *S* learns items in the order of their difficulty for *him*, the first n items that are learned by a given *S* in a list of a given length L ($L > n$) are likely to be easier for him than the fixed n items found in a list of length n, even though the lists are

designed to be homogeneous in difficulty for populations of *S*s.

An analysis of variance was also performed for all data in Table 3 pertaining to acquisition of complete lists (i.e., learning of the 4 pairs in a 4-pair list, of 8 pairs in an 8-pair list, etc.). All main effects were significant beyond the 1% level; *F* ratios with associated degrees of freedom were as follows: Presentation Rate, $F\ (2, 72) = 19.1$; Meaningfulness, $F\ (2, 72) = 105.2$; List Length, $F\ (3, 72) = 164.9$; Ability, $F\ (1, 72) = 10.2$. No interactions approach significance except Meaningfulness × Ability ($p = .10$) and Rate × Meaningfulness ($p = .16$). The main effects account for 83.8% of the total variance in the data, and to this degree it would be satisfactory to predict the observations by using the main effect means shown in Table 4 in the type of linear equation underlying the analysis of variance.

The significant effect obtained for rate of presentation requires comment in view of Bugelski's (1962) finding that within the limits of the conditions he used, total time for learning did not vary significantly even when rate of presentation varied. The present experiment was conducted before Bugelski's report appeared in the literature, and is not exactly comparable to Bugelski's: Bugelski used a constant stimulus-term time of 2 sec., exposed the response syllable for either 2, 4, 6, 8, or 15 additional sec., and allowed a 2-sec. interval between pairs. Thus, Bugelski's 2-sec. and 4-sec. response conditions amount, respectively, to total times of 6 and 8 sec. per pair. Further, Bugelski used college sophomore *S*s (in contrast to our tenth graders), set a criterion of two successive anticipations of a given pair, and gave no warm-up session. Nevertheless, our data for low-meaningfulness 8-pair lists at 4-sec. and 8-sec. rates of presentation are roughly comparable to Bugelski's data for his 8-pair nonsense-syllable lists at 6-sec. and 8-sec. exposure rates. Converting our mean $\log_e$ times of 5.93 and 6.13 for these conditions (from Table 3) to mean seconds per syllable, we obtain values of 48.2 and 52.5 sec., respectively, to be compared with Bugelski's mean values of 61.2 and 70.1 sec., respectively—reported by Bugelski as not significantly different. In view of the lack of comparability of procedures and analyses, the two sets of results are not seriously discrepant. But it will be noted that our range of presentation times (3, 4, and 8 sec.) has only a small overlap with the range of times used by Bugelski (6, 8, 12, and 19 sec. by a comparable reckoning). Bugelski's finding that total learning time is roughly constant applies, it would appear, only to relatively slow pres-

TABLE 4

Main Effect Means for Intact Lists (Log_e Sec.)

Presentation Rate (Sec.)	*M*	Meaningfulness	*M*	List Length	*M*	Ability	*M*
3	5.62	H	4.97	4	4.52	Fast	5.59
4	5.48	M	5.90	8	5.46	Slow	5.83
8	6.03	L	6.26	12	6.07		
				20	6.78		

Grand *M*: 5.71

entation rates. It is also probable that it applies more particularly to difficult, low-meaningfulness materials in which there are likely to be a large number of trials with the faster presentation rates. In our data, it is clear that high-meaningfulness materials are more efficiently presented with fast rates; a rate of 8 sec. tends to hold the learner back. For low-meaningfulness materials, on the other hand, there is some evidence of a curvilinear trend such that a 4-sec. rate is superior to either a 3-sec. rate or an 8-sec. rate. It was noted above that the Rate × Meaningfulness interaction approached significance ($p = .16$).

The demonstration that ability level was a significant main effect ($p < .01$) means merely that there is a significant correlation between an external measure of paired-associate learning ability (Part V of the MLAT) and ability manifested in the experimental task itself. This does not exclude the possibility that the experimental error ("within cells") includes ability variance not predicted by the external measure.

Fitting of data to Thurstone's rational equation for learning time as a function of list length.—Thurstone (1930, Equation 40) proposed that the relation between the number of items in a list and the total learning time is given by

$$T = \frac{c}{k} n\sqrt{n - a} \qquad [2]$$

where T = total learning time measured to any arbitrarily stipulated degree of perfection, n = number of items in the task, c = a constant depending on the complexity of the task and the degree of perfection required, k = S's learning ability, and a = "an adaptation constant of the subject . . . usually . . . his attention span." Assuming $a = 0$, taking logarithms, and introducing a constant s for the exponent of n (s being equal to 1.5 for Thurstone's equation with $a = 0$), we have the linear equation

$$\log T = s \log n + \log c - \log k \qquad [3]$$

The logarithmically plotted graph (Fig. 1) of the whole-list data from Table 3 suggests that these data could reasonably be fitted by Equation 3 using appropriate parameters for c and k. To test this observation, the total variance of the dependent variable expressed logarithmically was broken down into portions attributable to (*a*) meaningfulness, rate of presentation, and their interactions; (*b*) the regression on list length expressed logarithmically, separated into components attributable to a common slope for all conditions and to deviations from the common slope attributable to particular levels of the aforesaid main effects and their interactions, and (*c*) departures of array means from a common linear slope and from slopes for particular levels of the main effects and their interactions.[3] Ability level was not studied as an independent variable in this analysis; thus, the experimental error term includes variance that could be in part ascribed to ability level. As in the previous analysis, Meaningfulness and Rate of Presentation were highly significant sources of variance (the mean squares remaining the same and the size of the error term being only slightly different). For the Common Slope of log learning time as a linear function of log list length, $F(1, 108) = 460.9$, obviously highly significant. No other significant F ratios were found, from which

[3] The computational procedure is a special case of a highly generalized regression analysis procedure developed by Albert E. Beaton, Jr. (Beaton, 1964).

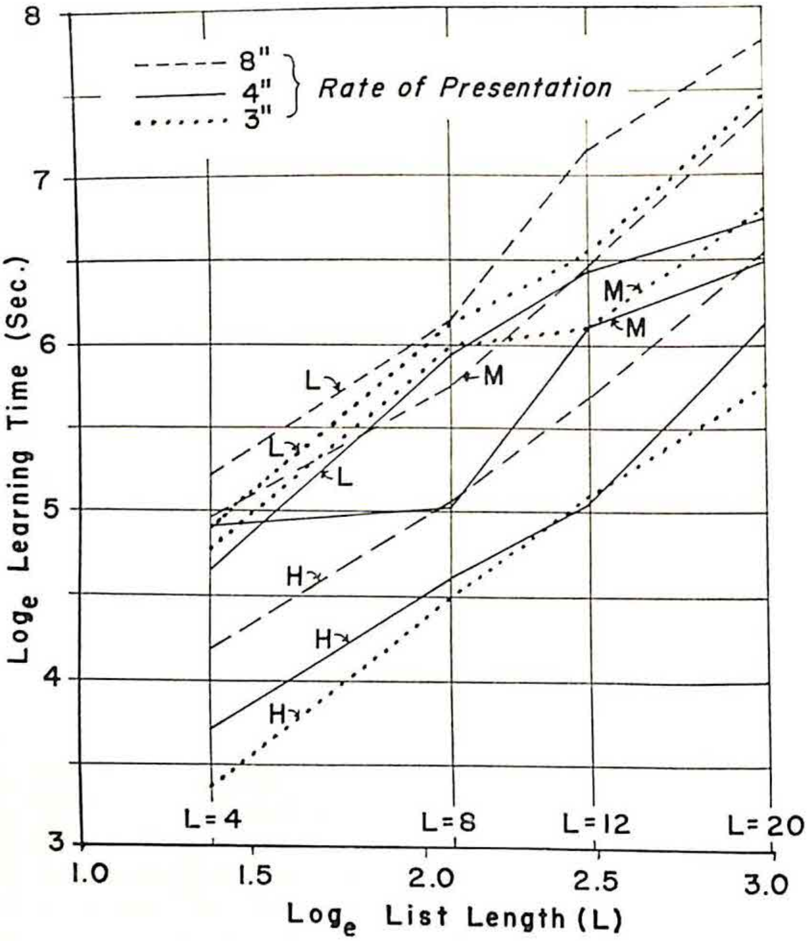


FIG. 1. Log learning times as a function of log list length, by meaningfulness (high, medium, low) and rate of presentation. (Each point plotted is a mean for four cases—two ability levels, two list forms.)

it may be concluded that (*a*) the nine curves shown in Fig. 1 do not depart significantly from linearity and (*b*) the slopes of the lines fitting these curves in a least-square sense do not depart significantly from a common value. Over the whole set of data, this regression slope is 1.4130, with 95% confidence limits 1.2876 to 1.5384 that obviously include Thurstone's rationally derived value of 1.5. The two main effects and the common linear slope account for 84.1% of the variance. It is thus possible to specify Equation 3 as

$$(\log T)_{\text{est}} = 1.4130 \log n + \log c_m + \log c_r + 2.5484 \quad [4]$$

where the values of $\log c_m$ and of $\log c_r$ are chosen depending upon the levels of meaningfulness and of rate of presentation, and 2.5484 is the intercept for the regression line averaged

TABLE 5

CONSTANTS FOR EQUATION 4

Meaningfulness	Log c_m	Rate (Sec.)	Log c_r
High	−.74	3	−.09
Medium	.19	4	−.23
Low	.55	8	.32

over all data. Log c_m and log c_r are actually the deviations of meaningfulness and rate array means (logs) from the grand mean (log), and are as shown in Table 5.

Role of ability.—Since ability was a significant main effect in Table 4, it is possible to supplement Equation 4 on the model of Equation 3 by adding a term for ability level, $(-\log k_a)$. In this case, log k_a for fast learners would be equal to .12, and for slow learners it would, of course, be $-.12$. But an even better prediction is obtained by using the actual scores on Part V of the MLAT and on the pretest. The correlations of these two measures with log learning times were, respectively, $-.13$ and $-.10$, and their intercorrelation was .24. It was found that an optimum composite of the two ability measures was a significant component of the variance of log learning times, $F\ (2, 136) = 26.9$, $p < .0005$. The introduction of continuous variables of ability into the prediction equation resulted in an overall slope of 1.4661, closer to Thurstone's rationally predicted value of 1.5.

It will be recalled, however, that in an analysis of variance cited earlier, Ability and Meaningfulness had a nearly significant interaction. If we

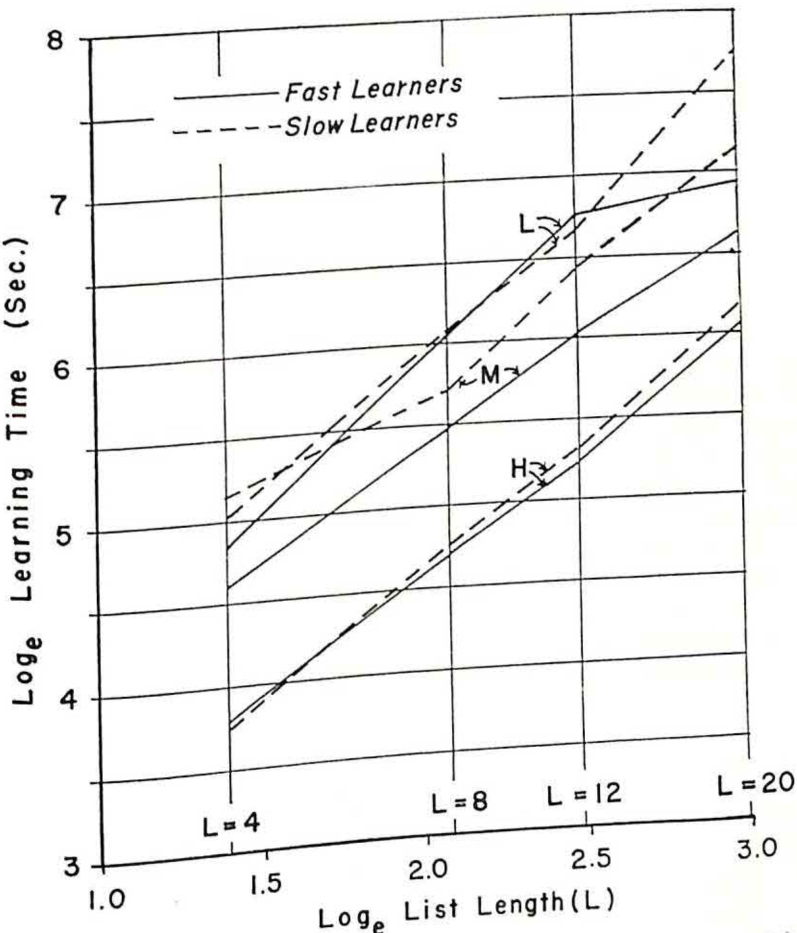


FIG. 2. Log learning times as a function of log list length, by meaningfulness (high, medium, low) and ability (fast, slow). (Each point plotted is a mean for six cases—three rates of presentation, two list forms.)

examine Fig. 2, which presents the data of Fig. 1 recombined for rates of presentation but broken down by Ability groups, we observe that Ability has a particularly noticeable effect for lists of Medium meaningfulness. A possible interpretation is that our paired-associate measure of learning ability (Part V of the MLAT) may be to some extent correlated with vocabulary knowledge, and if so, the low-frequency words in the medium-meaningfulness lists (see Table 1) would be more likely known to the fast learners than to the slow learners; for the latter, these low-frequency words would be virtually equivalent to nonsense syllables and thus rather similar to low-meaningfulness materials.

Discussion

The impetus for this study was the belief that information concerning acquisition rates for paired-associate materials under various conditions would be of use in planning efficiently programed instructional materials.

One of the obvious conclusions to be drawn from the experiment is that depending upon the nature of the materials, the conditions of training, and the ability of *S* there are certain limits to the rate at which new responses can be acquired. The experiment confirms Thurstone's (1930) prediction that as the number of items to be learned increases, there is an increase in the amount of time that must be spent per item. For if we divide Equation 2 through by n in order to get t, the learning time per item, we find that t increases approximately as the square root of the number of items. The only hope of getting people to learn more efficiently is to vary the conditions of training (including, presumably, variations in incentive, which were not explicitly studied here). The immediate suggestion coming from the experiment is that the optimal rate of presentation in a paired-associate experiment of this type is approximately 4 sec., as contrasted to 3 or 8 sec.

A major insight that comes from this study and that was somewhat unanticipated was that items appear to be acquired one by one and at a rate that is relatively independent of the total number of items presented in a given trial or session. For example, it has been shown in this study that it takes about the same amount of exposure and response time for an *S* to acquire 4 items regardless of whether they stand alone or are imbedded in a list of 20 items. It is possible that Thurstone's square-root law is limited to situations in which the whole of a list is presented in each trial and that it would not apply in the same way to some other organization of the learning task, e.g., a progressive part method. This problem has not been adequately investigated for paired-associate learning.

It would be desirable to obtain data for lists much longer than 20 pairs in order to verify the postulated equation. Thurstone (1930) has already discussed the application of his equation to several sets of data involving long lists of nonsense syllables or digits to be learned serially. There is some credibility in extrapolating the present data, however, when we try to predict Thorndike's (1908) finding that on the average, a group of students learned the English meanings of 1,030 German words in 30 hr. Assuming that learning of English meanings of German words is a task equivalent in difficulty to our medium-meaningfulness level, and that Thorndike's students tended to select the most efficient presentation rate for this level of difficulty (4 sec. per pair), we may apply Equation 4 and estimate that on this basis 1,030 pairs would require about 222,100 sec. or about 61.6 hr., a value which is in the same decimal order of magnitude as Thorndike's finding. (Actually, Thorndike's students spent up to an additional 8 hr. testing themselves, and undoubtedly some learning occurred here.)

References

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams.

Psychol. Monogr., 1960, **74** (10, Whole No. 497).

BEATON, A. E., JR. The use of special matrix operators in statistical calculus. Unpublished doctoral dissertation, Harvard University, 1964.

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 409–412.

CARROLL, J. B., & SAPON, S. M. *Modern language aptitude test.* New York: The Psychological Corporation, 1958.

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m') in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

HOVLAND, C. I. Experimental studies in rote-learning theory: VII. Distributed practice of paired associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.

KJELDERGAARD, P. M. Predicting paired-associate learning speed. *Psychol. Rep.*, 1962, **11**, 353–354. (a)

KJELDERGAARD, P. M. Variables affecting recall in one trial learning. Paper read at Eastern Psychological Association, Atlantic City, April 1962. (b)

THORNDIKE, E. L. Memory for paired-associates. *Psychol. Rev.*, 1908, **15**, 122–138.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

THURSTONE, L. L. The relation between learning time and length of task. *Psychol. Rev.*, 1930, **37**, 44–58.

WILLIAMS, J. P. Supplementary report: A selection artifact in Rock's study of the role of repetition. *J. exp. Psychol.*, 1961, **62**, 627–628.

(Received March 20, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 554–563

ROLE OF APPARENT SLANT IN SHAPE JUDGMENTS[1]

WILMA A. WINNICK AND ILANA ROGOFF

Queens College of the City University of New York

2 parallel experiments are reported bearing on the shape-slant invariance hypothesis. Apparent-objective slant scales were 1st determined for 4 rectangles, 2 trapezoids, a random shape, and an ellipse. Apparent slant was found to be less than objective slant at 10°, 20°, and 30°, and to be greater at 60° and 80°; accuracy was achieved at close to 40°. Shape judgments were then measured for the same forms set at the same angles. The obtained increases in the quantity a − p as a function of slant agreed with predictions from the slant scales. Not predicted was the finding of overconstancy at 10° and 20°.

The experiments to be reported comprise an attempt to draw a research parallel between size constancy and shape constancy, a parallel which has been pursued on the theoretical plane by Koffka's (1935) treatment of the constancies. Apparently, size constancy has challenged *E*s much more than has shape constancy, and a wealth of experimentation on its parameters has been forthcoming (Woodworth & Schlosberg, 1954). Shape, on the other hand, has been shown in a recent review (Epstein & Park, 1963) to be lacking in ". . . well established functional relationships."

With size constancy apparently dependent upon perception of distance, shape constancy would seem to be dependent upon perception of the angle of orientation of the viewed object. Beck and Gibson (1955) have formalized this relationship as "the shape-slant invariance hypothesis": "A retinal projection of a given form determines a unique relation of apparent shape to apparent slant [p. 126]." Although a few *E*s have studied judgments of slant in shape-constancy experiments (e.g., Stavrianos, 1945), and others have studied cues to slant judgments (e.g., Smith, 1956), no experiment to the writers' knowledge has attempted to measure a scale of slant judgments paralleling Gilinsky's (1951) determination of a scale of distance judgments and further to relate such a scale to shape constancy as a function of slant. The value of this scale is indicated by the statement that "the adequacy of the (invariance) hypothesis depends on the possibility that the various factors whose influence on shape constancy has been demonstrated may be shown to affect perceived slant [Epstein & Park, 1963, p. 283]."

The first experiment reported, accordingly, undertook the determination of the nature of the relationship between apparent and objective slant as a function of conditions of observation and characteristics of the form whose slant was to be judged. These determinations were followed, in the second experiment reported, by measurements of shape judgments for the same stimuli at the same angles of slant.

EXPERIMENT I

Method

Experimental design.—Four parallel studies were carried out; the first two measured ap-

[1] The slant experiments are based upon a thesis submitted by the second author to the Graduate Division of Queens College of the City University of New York in partial fulfillment of the requirements for the degree of Master of Arts. This study was partially supported by a grant from the Graduate Division of the City University of New York.

parent-objective slant scales for four rectangles at 20°, 40°, 60°, and 80°, the first under restricted conditions, the second with full view; the third study measured the scale for two trapezoids, an ellipse, and a random shape at the same angles; the fourth measured the scale for one rectangle set at 10°, 20°, 30°, and 40°. In each of these independent studies, all *Ss* judged each form used at each of the angles studied.

Apparatus.—The *Ss* were required to judge, using binocular vision, the angle of inclination of a single black figure set 2 ft. away in the center of a gray experimental chamber of 16.25 in. in width, 21 in. in height, and 2 ft. in depth. In the first study, the figures were viewed through a 1-in. opening which permitted a view of the middle part of the rectangles only. This was judged to be similar to Lichte's (1952) experimental situation, in which judgments of shape were based on the width of the rectangles only. In the other studies removal of the top of the chamber provided *Ss* with a view of the tops of the stimulus figures, as well as other parts of the chamber.

Subjects.—The 64 volunteer *Ss*, 16 in each experiment, were university undergraduates ranging in age from 19 to 27 yr. None had serious visual defects, and 30 wore corrective lenses at all times (and for this experiment).

Procedure.—The method used in each of the four studies was a modification of the method of average error, in which judgments were made about a single stimulus. On each trial, the task of *S* was to indicate when, as *E* varied the slant of the figure by rotating it to *S*'s right, it appeared to be inclined at the angle called for (20°, 40°, 60°, and 80° in the first three studies and 10°, 20°, 30°, and 40° in the fourth). To insure uniform knowledge of angles, each *S* was given a linear representation of the angles used.

The session began with an explanation of the task and apparatus and two trials to test *S*'s understanding of the instructions. In each experiment each of the four forms was presented with four angles for eight trials in both ascending and descending determinations. Blocks of trials were arranged by figure and varied as follows: ABCD, CBDA, ADBC, DCBA. The within-blocks trials were arranged randomly for angles for both ascending and descending trials.

All manipulations of forms were begun randomly either from the frontal-parallel plane (ascending) or at right angles (descending). The form was slowly turned by *E* in the vertical plane until *S* reported that it was inclined on his right side at the required angle from the frontal-parallel plane. The *E* recorded the exact setting at which *S* judged the inclination to be at the angle called for (from a protractor located at the base of the dowel supporting the figure).

Stimuli.—The forms used were, for the first two studies, rectangles all 5 in. in height but varying in width, as follows: (A) 3.25, (B) 3.75, (C) 4.25, and (D) 4.75 in. These rectangles correspond to those used by Lichte (1952) in his study of shape constancy.

For the third study, the forms were two trapezoids, one ellipse, and one random shape. The two trapezoids had identical dimensions (5 in. in width and 10 and 7.5 in. in height), one with the obtuse angle on *S*'s left (C) and the other on *S*'s right (D). The trapezoids were such that their visual (geometric) projection was equal to that of a rectangle viewed at a tilt of 26° (tilt to the left for C and to the right for D). In the ellipse, the height of the vertical diameter was 11 in. and of the horizontal diameter, 5 in. The random shape used was Shape B from Borresen and Lichte (1962).

Since no effect of rectangle width had been found in the first two studies, the fourth used only the rectangle of 4.75 in. width.

Results and Discussion

Table 1 summarizes the mean settings obtained (objective slant) for the angles called for (apparent slant), and Table 2 summarizes the results of the analysis of variance for the four studies.

TABLE 1

MEAN SETTINGS IN DEGREES TO MATCH STANDARD ANGLES IN THE FOUR STUDIES OF EXP. I

Rectangles	20°	40°	60°	80°
Study I, Restricted View				
A. 3.25 in. width	25.9	39.6	55.7	72.2
B. 3.75 in.	23.6	38.0	51.0	70.0
C. 4.25 in.	24.1	37.1	53.2	70.9
D. 4.75 in.	23.2	37.8	53.1	70.5
Study II, Full View				
A. 3.25 in.	26.0	39.4	53.1	70.5
B. 3.75 in.	25.5	39.9	52.7	68.4
C. 4.25 in.	26.2	29.3	52.3	68.4
D. 4.75 in.	24.8	39.9	53.3	68.8
Study III, Full View				
A. Ellipse	23.7	37.0	53.9	70.8
B. Random	22.7	37.3	54.6	71.5
C. Trapezoid	22.0	34.7	54.0	71.3
D. Trapezoid	24.0	38.7	57.4	74.3
Study IV	10°	20°	30°	40°
4.75 in.	17.5	24.1	32.4	38.5

TABLE 2

ANALYSIS OF VARIANCE FOR MEAN JUDGMENTS IN THE FOUR STUDIES OF EXP. I

Source	*df*	Study I *MS*	Study I *F*	Study II *MS*	Study II *F*	Study III *MS*	Study III *F*	Study IV *df*	Study IV *MS*	Study IV *F*
Ss	15	134.5		73.5		105.6		15	45.32	
Angles (A)	3	25,776.20	7.99**	21,979.0	792.6**	29,829.0	926.4**	3	1,370.32	835.56**
Forms (F)	3	94.2		6.6		202.6	19.3**			
Ss×F	45	238.8		6.8		10.6				
Ss×A	45	322.5		27.7		32.2				
A×F	9	7.2		6.8		30.2	16.1**	45	1.64	
Ss×A×F	135	126.7		4.5		1.9				
Total	255							63		

** $p < .01$.

The data obtained in the first two studies (restricted and full view of the four rectangles) are highly similar, indicating the use of similar cues in the two situations. Both analyses of variance obtained a significant F only for angles, with F values for rectangles and interaction of rectangles and angles falling short of significance. (The statistical design does not permit investigations of the other interactions nor of the triple interaction.)

The graph relating apparent to objective slant for the 4.75-in. rectangle under full viewing conditions is shown in Fig. 1 and will be considered to represent the apparent-objective slant scale for rectangles, since data

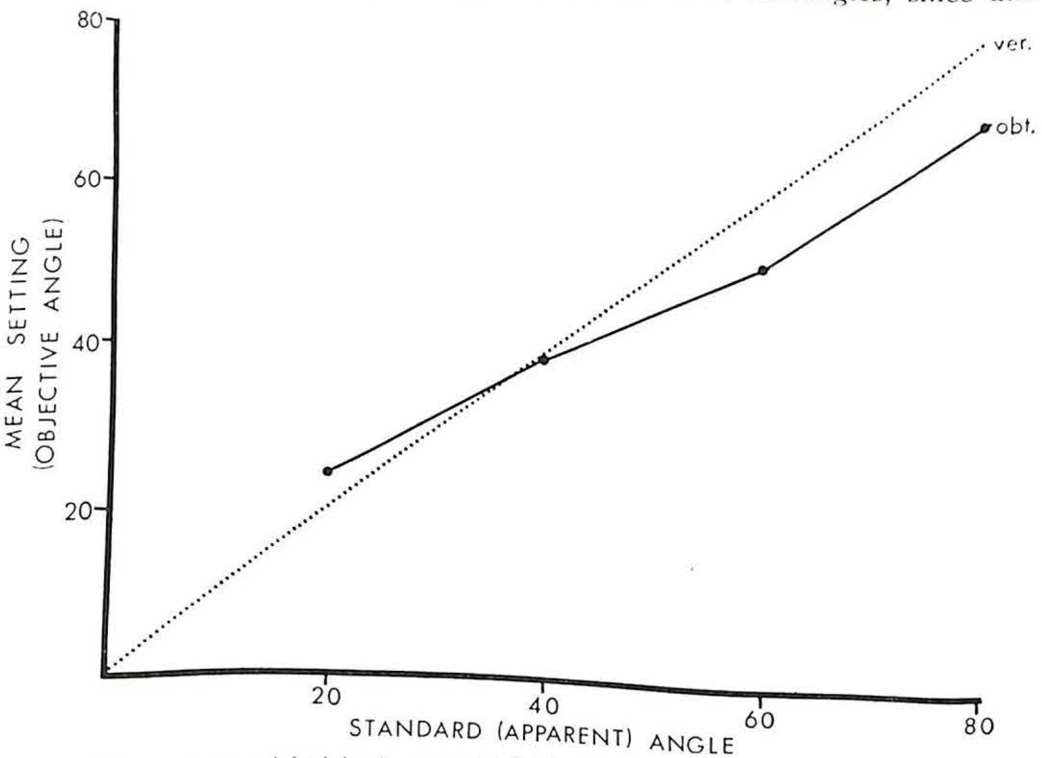


FIG. 1. Mean settings (obt.) in degrees of 4.75-in. rectangle for each of the four standard angles under full view conditions, compared to veridical judgments (ver.).

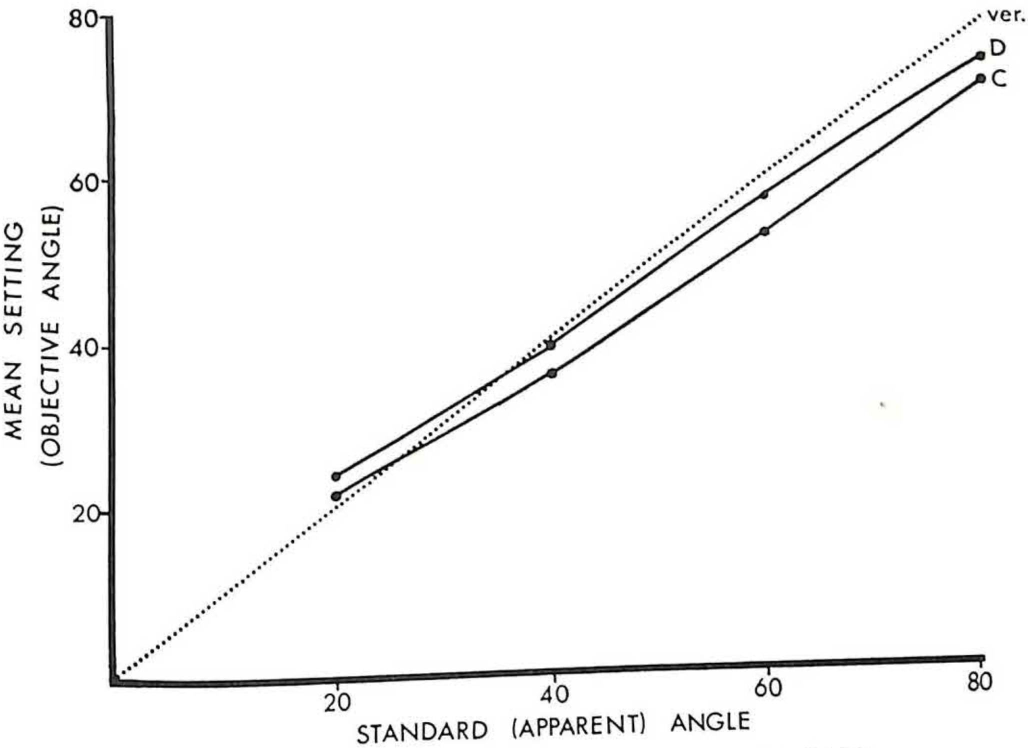


FIG. 2. Mean settings in degrees for the two trapezoids (C, D) compared to veridical judgments (ver.).

from the first two studies were close to identity and showed no effect of rectangles. Statistical comparisons of obtained with veridical points yielded the following t (15) values: 5.5 for 20°, 6.6 for 60°, and 9.8 for 80° ($p < .01$ for all). These results indicate apparent slant to be significantly less than objective (a positive constant error) at 20° and to be greater (a negative constant error) at 60° and 80°.

Where nonrectangular forms were used, in the third study, the results of the analysis of variance (Table 2) showed a significant F for forms ($p < .01$) and for Forms × Angles interaction ($p < .01$), as well as for angles ($p < .01$). Although no differences were found in the scales for the ellipse and the random shape, a clear-cut difference in settings for the two trapezoids is evident from the graph in Fig. 2. Comparisons of obtained settings with veridical points found t (15) values for Trapezoid C of 2.7 at 20° ($p < .05$), 3.8 at 60° ($p < .01$), and 5.7 at 80° ($p < .01$); for Trapezoid D, obtained t values were 4.2 at 20° ($p < .01$), 2.0 at 60° ($p < .05$), and 4.8 at 80° ($p < .01$). Furthermore, a comparison of mean pooled judgments for the two trapezoids yielded a t (30) = 8.6 ($p < .01$). This contrast in settings of the two trapezoids lends some support to a Pragnanz view of shape constancy (Epstein & Park, 1963, p. 275).

The fourth study demonstrated, in its extension of the scale toward zero, that the positive error found at 20° occurred also at other small angles. The resultant mean judgments for the one rectangle (4.75 in.) used are shown in Table 1, the analysis of variance in Table 2, and the apparent-

objective slant scale for these angles in Fig. 3. The analysis of variance found the F value for angles to be significant at better than the .01 level. Comparisons obtained with veridical judgments yielded the following t (15) values: 15.0 for 10° ($p < .01$), 4.1 for 20° ($p < .01$), and 2.7 for 30° ($p < .05$).

Since no previous study has investigated the relationship between apparent and objective slant over as full a range of angles, comparisons of the data must be made with studies using smaller segments of the arc. Studies by Clark and his associates (Clark, 1953; Clark, Smith, & Rabe, 1955, 1956a, 1956b) have been directed toward identification of monocular cues to slant and have found apparent slant to be less than objective slant for angles of 53° (Clark, 1953), of 0°, 20°, and 40° (Clark et al., 1955), and of 40° (Clark et al., 1956a, 1956b). In Smith's (1956) study of monocular judgments of six slants (10°, 20°, 30°, 40°, 50°, and 60°) perceived slants were found, in partial agreement with the present study, to be less than objective slant for all angles.

Applying these data directly to the invariance hypothesis, one may predict that, if constancy of shape is dependent upon accuracy of slant judgments, shape estimates for rectangles and ellipses should be most accurate from 35° to 40° (from 25° to 30° for Trapezoid C, 38° to 40° for Trapezoid D) and should fall off for smaller and for larger angles. Further, if the slant of an object is underestimated (as for the smaller angles and for Trapezoid C), its perceived shape should be close to projected (retinal) shape. On the other hand,

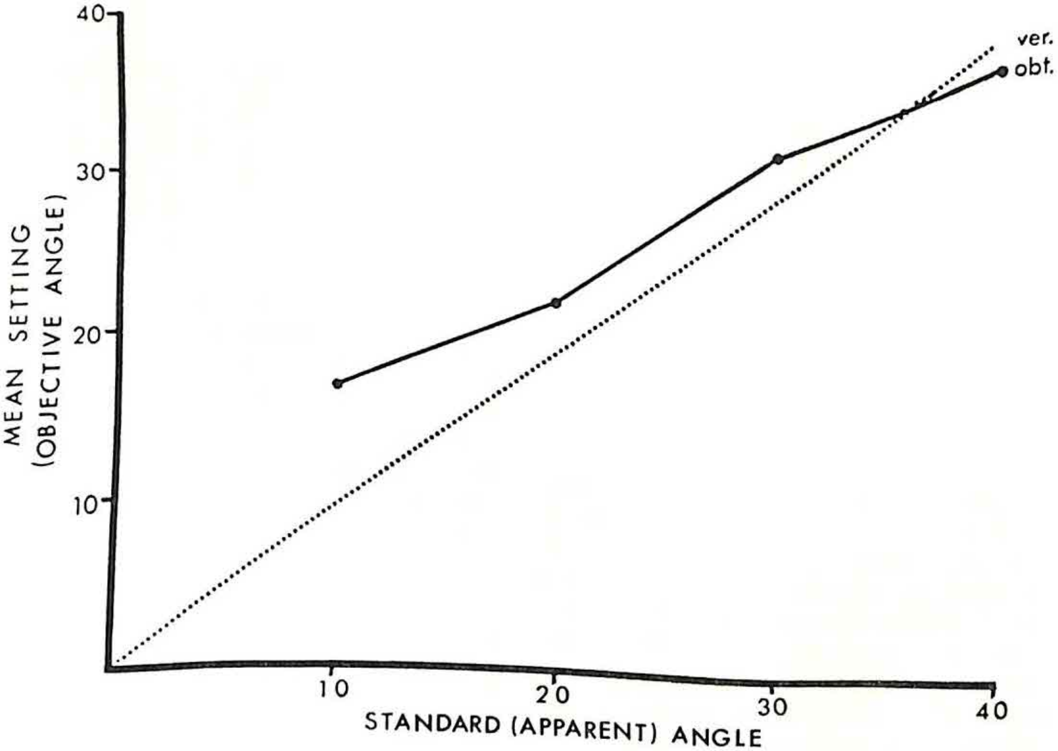


FIG. 3. Mean settings in degrees for the 4.75-in. rectangle at the four standard angles in the fourth study, compared to veridical judgments.

TABLE 3

MEAN WIDTH SETTINGS IN IN. TO MATCH THE SLANTED FORMS IN EXP. II

	10°	20°	30°	40°	60°	80°
4.75-in. rectangle	4.86	4.93	4.79	4.37	4.34	3.34
5.00-in. Trapezoid C	5.05	4.90	4.91	4.81	4.33	3.56
5.00-in. Trapezoid D	5.26	5.13	5.20	4.96	4.62	3.53
5.00-in. ellipse	5.03	5.10	5.00	4.95	4.78	4.30

where slant is overestimated (as for the larger angles and for Trapezoid D), perceived shape should be far from projected shape.

The relevant constancy measure into which these predictions might be transposed is given by the quantity $a - p$, where a is apparent shape and p is projected (retinal) shape. Using this formula as the basis for predictions, data should reveal an increase in this value with increasing angles. Furthermore, the quantity $a - p$ should be small for Trapezoid C and large for Trapezoid D, in comparison with its value for rectangular forms.

EXPERIMENT II

To provide the data needed to assess the predictions of shape as a function of slant, based on the shape-slant invariance view, two studies were undertaken measuring shape judgments for the forms used in the previous experiment set at the same angles of inclination in the same apparatus.

Method

Apparatus and procedure.—Standard stimuli (the 4.75-in. rectangle and the 5.00-in. trapezoids in the first study; the 5.00-in. ellipse in the second) were presented at 10°, 20°, 30°, 40°, 60°, and 80° in the chamber already described. Both studies presented comparison stimuli in the frontal-parallel plane, but the methods of matching had to be somewhat different, as dictated by the nature of the stimuli.

For measuring constancy for the rectangle and the two trapezoids, a comparison stimulus was presented equal in height to the standard and of width that could be varied to match the width of the standard in accordance with the method of average error. More specifically, the apparatus consisted of a central black Masonite rectangle set immediately behind and framed by a free-standing gray Masonite rectangle 16.25 × 21 in. This gray rectangle, equivalent in outside dimensions to the experimental chamber, was set at the same distance from *S* as was the

TABLE 4

ANALYSIS OF VARIANCE OF MEAN SETTINGS OBTAINED IN EXP. II

Source	Rectangle and Trapezoids			Ellipse		
	df	*MS*	*F*	*df*	*MS*	*F*
Ss (*Ss*)	14	.43		14	.25	
Angles (A)	5	15.96	37.12**	5	1.30	7.22**
Forms (F)	2	1.82	4.23*			
Ss × A	70	.60		70	.18	
Ss × F	28	.36				
A × F	10	.43				
Ss × A × F	140					
Total	269			89		

* $p < .05$.
** $p < .01$.

standard rectangle within the chamber. The black variable rectangle, centered within a cutout from the large gray rectangle, could be varied in width from 2.0 in. to 7.5 in. by moving on a pulley system a vertical piece equivalent in height and general appearance to the gray framing rectangle and 6.0 in. in width. For each figure at each angle, *E* varied the width of the comparison stimulus until *S* judged its width equal to that of the standard; two determinations, one ascending and one descending, were made for each *S* for each form at each angle. A randomized arrangement of the three forms and six angles was employed across *S*s.

In the second shape study, comparisons were made of the standard ellipse (11 × 5 in.) with variable ellipses all 11 in. in major axis and with minor axes varying in units of .25 in. from 2.0 to 6.0 in. The variable ellipses were presented one by one in randomized arrangement, in accordance with the method of constant stimuli. Two determinations were made for each *S* at each of the six angles of slant. Order of presentation of angles was varied across *S*s.

Subjects.—Thirty university undergraduates served as *S*s, 15 in each study. All *S*s had had previous experience in making the judgments called for; in addition, half had served as *S*s in the slant experiments.

Results and Discussion

Table 3 summarizes the data obtained in the two shape-judgment studies, and Table 4, the analyses of variance. In both, the *F* values for angles were significant at the .01 level, and for the first study, the *F* value for forms reached significance at better than the .05 level.

The relationship between apparent shape and angle of inclination is shown by the graph in Fig. 4 for the rectangle and in Fig. 5 for the two trapezoids and the ellipse. These graphs also plot object shape and projected shape, permitting estimates of the value $a - p$. All of these functions show similar relationships—overconstancy at 10° and 20°, near object matches at close to 30°, and matches falling away from projected

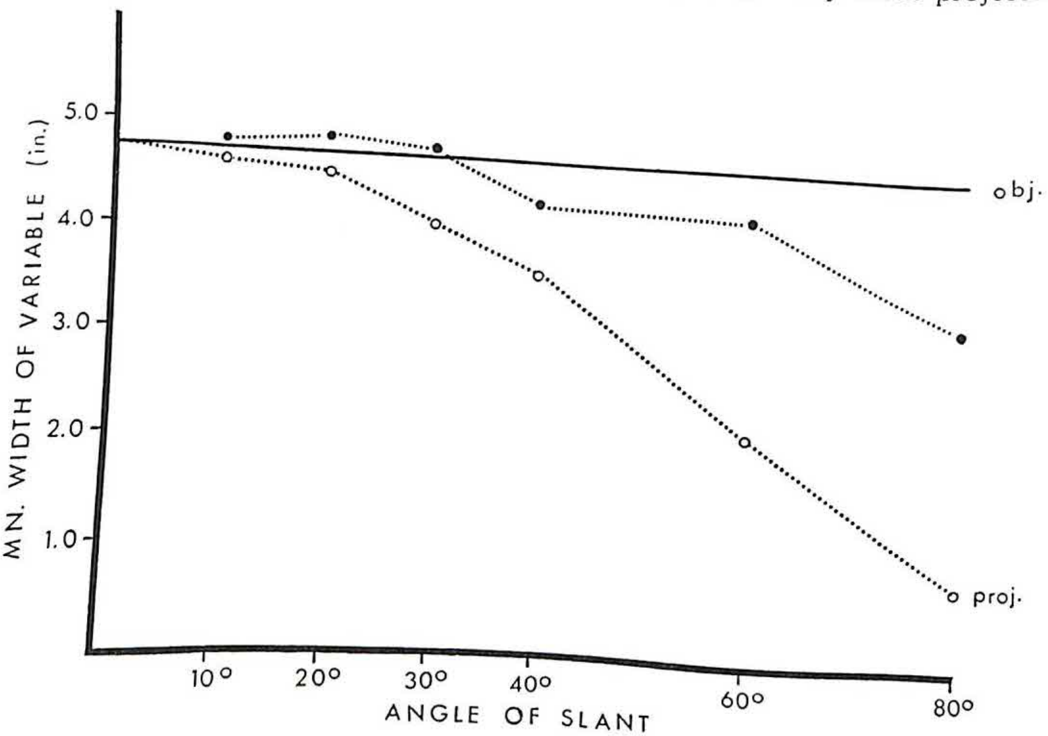


FIG. 4. Apparent shape as a function of angle of inclination for the 4.75-in. rectangle, compared to object (obj.) and projected (proj.) shape.

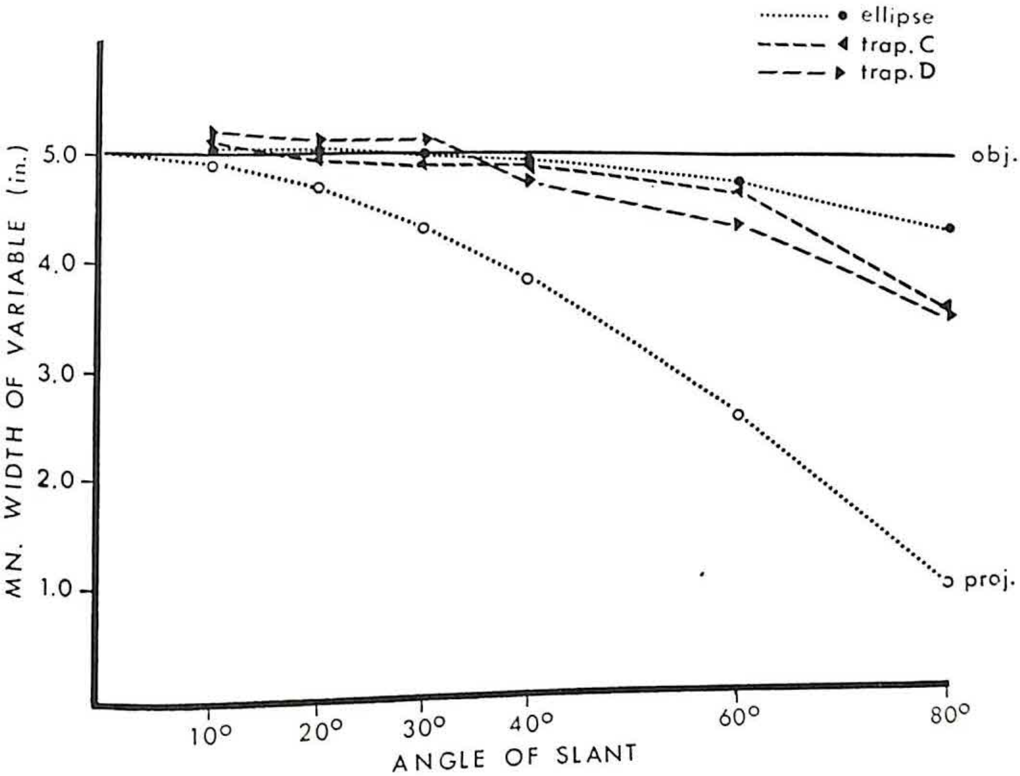


FIG. 5. Apparent shape as a function of angle of inclination for the 5.00-in. trapezoids and the ellipse, compared to object (obj.) and projected (proj.) shape.

width at the larger angles. For all forms, the quantity $a - p$ may be seen by inspection to increase as a function of angle, and Fig. 6 shows a plot of this relationship.

The data obtained for the ellipse may be compared with results obtained by Nellis (1958) in an experiment measuring shape judgments for ellipses set at angles of 15°, 30°, 45°, 60°, and 75°. Using a somewhat different method, Nellis' obtained functions relating judgments to angle of inclination are similar to those in Fig. 5; she found *S*s making their closest matches to object size at close to 30° of slant, with overestimation at the smaller angles and underestimation at larger angles, as in the present study. She (Nellis, 1958, p. 49) computed the value a/p and found an increase in this quantity for a 5.5 ellipse from 1.06 at 15° to 2.82 at 75°. Converting the present data into a/p values, there is a comparable increase for the 5.0-in. ellipse used from 1.02 at 10° to 4.23 at 80°.

General Discussion

Confronting the shape-constancy data obtained in the second experiment with the predictions based on the scales of apparent-objective slant from the first experiment, the data lend some support to the shape-slant invariance hypothesis. Predictions as to the locus of accuracy of shape matches were not confirmed with precision; closest match to object size was made for the rectangle and the ellipse at close to 30°; i.e., at a smaller setting than predicted. The two trapezoids, equivalent to rectangles slanted in opposite directions, showed some contrast in shape judgments, though not of as great magnitude as for slant estimates.

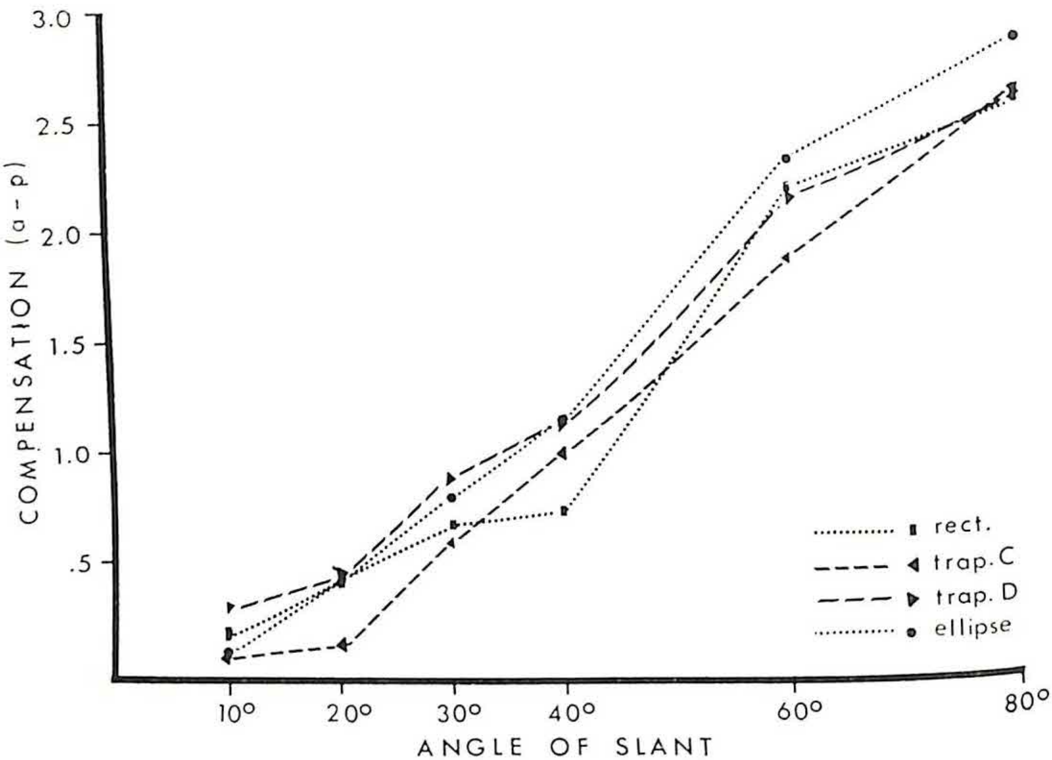


FIG. 6. Compensation as a function of angle of inclination for the rectangle, two trapezoids, and the ellipse.

The data generally support the prediction that, along with underestimations of slant found at the smaller angles, matches should be close to projected shape, and that overestimations of slant at larger angles should result in settings progressively farther away from projected shape. What could not be predicted from the apparent-objective slant scale as such was the finding of overconstancy at 10° and at 20°. These angles were those at which *S*s had underestimated slant, and it was predicted that shape judgments should be close to projected shape; since, however, at these small angles, projected shape is itself close to object shape, judgments must also be close to object shape. If we view *S* as aware of the inclination of the form and therefore tending to compensate for the angle of slant, then his attempt to do this may cause his judgment, in retreating from projected shape, to go beyond object shape. Nellis (1958) has applied this kind of explanation to her findings of overconstancy at smaller angles:

> Consider the larger angles of tilt where compensation is manifested by the fact that the match lies between projected image-size and object-size. If there is compensation with small angles of tilt, then perceived size which is consistently greater than projected image-size also yields a match greater than actual size, since projected image-size is so close to object size [p. 87].

This compensatory tendency, then, appears to be an effect additional to the shape-slant relationship and sometimes opposing it. The shape-at-a-slant influence directs perception toward retinal shape; on the other hand, awareness of the slant itself may cause a movement of the perception toward the real object. For most angles of orientation, the result of these two processes is a compromise lying somewhere between retinal shape, as determined by the shape-slant invariance relationship, and the object shape,

as dictated by the compensatory tendencies within *S*. In the case of the smaller angles, however, since the shape-slant effect is so slight, the compensation seems dominant.

This two-aspect view is, of course, in different terms what Koffka (1935) has asserted to be the case:

According to our assumption a retinal shape sets up forces to produce a similar psychophysical shape. These forces come into conflict with the stress in the field when the plane in which the shape appears is nonnormal. Owing to this stress the retinal shape is changed into another shape more like the real one [p. 233].

The view seems similar also to one that Beck and Gibson (1955) have attributed to Helmholtz:

A follower of Helmholtz . . . would argue that the organism discovers in the course of time that the cues for the slant of an object vary concomitantly with the retinal projected form of the object. Only then is he in a position to infer the true shape of the object from the retinal shape by combining this information with the cues for slant [p. 127].

REFERENCES

BECK, J., & GIBSON, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50**, 125–133.

BORRESEN, C. R., & LICHTE, W. H. Shape constancy: Dependence upon stimulus familiarity. *J. exp. Psychol.*, 1962, **63**, 91–97.

CLARK, W. C. Exposure time and the perception of slant. Paper read at the Canadian Psychological Association, Kingston, Ontario, 1953.

CLARK, W. C., SMITH, A. H., & RABE, A. Retinal gradients of outline as a stimulus for slant. *Canad. J. Psychol.*, 1955, **9**, 247–253.

CLARK, W. C., SMITH, A. H., & RABE, A. The interaction of surface texture, outline gradient, and ground in the perception of slant. *Canad. J. Psychol.*, 1956, **10**, 1–9. (a)

CLARK, W. C., SMITH, A. H., & RABE, A. Retinal gradients of outline distortion and binocular disparity as stimuli for slant. *Canad. J. Psychol.*, 1956, **10**, 77–81. (b)

EPSTEIN, W., & PARK, J. Shape constancy: Functional relationships and theoretical formulations. *Psychol. Bull.*, 1963, **60**, 265–288.

GILINSKY, A. S. Perceived size and distance in visual space. *Psychol. Rev.*, 1951, **58**, 460–482.

KOFFKA, K. *Principles of Gestalt psychology.* N. Y.: Harcourt Brace, 1935.

LICHTE, W. H. Shape constancy: Dependence upon angle of rotation: Individual differences. *J. exp. Psychol.*, 1952, **43**, 49–57.

NELLIS, B. S. Effects of object- and background-tilt on perception of form. Unpublished doctoral dissertation, University of Texas, 1958.

SMITH, A. H. Gradients of outline convergence and distortion as stimuli for slant. *Canad. J. Psychol.*, 1956, **10**, 211–218.

STAVRIANOS, B. K. The relation of shape perception to explicit judgments of inclination. *Arch. Psychol., N. Y.*, 1945, No. 296.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received March 5, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 564–578

NEUTRALIZATION OF STIMULUS BIAS IN THE RATING OF GRAYS [1]

IRWIN POLLACK

Mental Health Research Institute, University of Michigan

Techniques for iterative experimentation, whereby the results of the 1st experiment modify the conditions of the 2nd, and the results of the 2nd in turn modify the conditions of the 3rd, etc. were examined for neutralization of the effects of stimulus bias in the psychophysical scaling of gray papers. The scaling methods of category estimation, category sorting, magnitude selection, and ideal selection were employed. The techniques were tested by deliberately introducing stimulus bias with sets of stimuli, sharply weighted in opposite directions, and determining whether the iterative procedure would yield a single empirical scaling from both sets of stimuli. In most instances, the techniques of iterative experimentation were effective in getting rid of the effects of stimulus bias quickly. Exceptions to the general finding are sorting and selection methods with extremely weighted or extremely narrow initial sets of stimuli.

An experiment by Garner (1954) serves as an important reminder that the validity of an apparently straightforward psychophysical experiment may be ruthlessly destroyed by seemingly innocent decisions made by *E* in the initial stages of investigation. Garner showed that the sound level required for half loudness could be entirely determined by the range of sound levels arbitrarily selected by *E* in the method of constants. Specifically, Garner exposed three groups of *S*s to pairs of tones. The *S* was instructed to indicate whether the second tone was more than, or less than, twice as loud as the first tone. The first member of the pair was constant at 90 db. The second member of the pair varied from 55 to 65 db. for one group, from 65 to 75 db. for another, and from 75 to 85 db. for a third group of *S*s. The mean and median levels required for half loudness were indistinguishable from the midpoints of the stimulus ranges. The means were: 60 db. for the first group, 70 db. for the second, and 80 db. for the third. The final result, then, was indistinguishable from a result completely determined by the stimuli arbitrarily chosen by *E*.

Stimulus bias will be said to be in evidence whenever the result of a psychophysical study is strongly determined by the particular stimuli arbitrarily selected by *E*. Garner's experiment is an excellent, although extreme, demonstration of stimulus bias.

The aim of the present paper is to describe, and test, a set of procedures for getting rid of—or, more precisely, for neutralizing—stimulus bias in psychophysical experiments. Stated differently, we are concerned with procedures which protect data from the seemingly innocent decisions made by

[1] This research was carried out at the Applied Psychology Research Unit, Cambridge, England, while the writer was a National Science Foundation Postdoctoral Research Fellow (1960–61). The writer is especially indebted to Donald Broadbent, Director, APRU, for providing a hospitable and encouraging research environment and to Christopher Poulton for long discussions and for arrangements in testing Cambridge University students. The writer is indebted to S. S. Stevens for discussions of early drafts of this paper.

E in his arbitrary selection of stimuli preparatory to an experiment. And, since rating scales are particularly vulnerable to the effects of stimulus bias (J. C. Stevens, 1958), primary emphasis will be placed upon getting rid of stimulus bias in rating scales. The aim is, thus, in sympathy with recent studies in the "social psychology of the psychological experiment [Rosenthal, 1963]."

One set of procedures which may neutralize stimulus bias is iterative (Stevens, 1955). Iterative procedures —in which the results of the first study modify the conditions for the second, the results of which, in turn, modify the conditions for the third, etc.—approach a final result through a series of successive approximations. On several occasions, Stevens and his co-workers (Poulton & Stevens, 1955; J. C. Stevens, 1958; Stevens, 1955, 1957; Stevens & Galanter, 1957; Stevens & Poulton, 1956; Stevens, Rogers, & Herrnstein, 1955) have pointed out the potential of iterative experimental procedures for obtaining unbiased psychophysical functions, although they rarely carried through the iterative procedures in a series of systematic investigations. One exception was the use of an iterative procedure for the unbiased averaging of data (Stevens, 1955). More recently, Smith (1961) has advocated the use of iterative procedures for stimulus programing in order to increase the efficiency of psychophysical experimentation. An excellent example of Smith's thesis is the Bekesy technique, as applied to human (Hirsh, 1962) and to animal (Blough, 1955) threshold testing.

While major efforts are presently directed toward the isolation of contextual factors in psychological experimentation (Helson, 1964; Johnson, 1955; Parducci, 1963; Rosenthal, 1963), the essential mode of operation of stimulus bias was described by Stevens and Galanter (1957). Typically, *S* expects to employ all of the available ratings in a rating experiment equally often (Arons & Irwin, 1932). When sets of stimuli are employed which substantially unbalance the equal frequency of ratings, *S* must wrestle with a conflict between performing the rating task at hand and his expectations of a uniform distribution of ratings. One method for avoiding this confiict is to tailor the stimuli, through iterative procedures, to conform with *S*'s expectations of a nearly uniform distribution of response categories. The following techniques attempt to provide a more nearly uniform distribution of response categories, and thereby, achieve rating scales which are relatively independent of the initial stimuli arbitrarily selected by *E*.

ITERATIVE TECHNIQUES

Technique of adjusted spacing.—This technique leads to additional stimuli in the stimulus region where sensitivity is high, i.e., where the slope of the category function is steep; and to fewer stimuli in the stimulus region where sensitivity is low, i.e., where the slope of the category function is shallow. This technique, along with the other techniques, makes the weak assumption that stimulus objects may be arranged to provide a monotonic category scaling. The technique also assumes that: intermediate stimuli may be interpolated along the stimulus scale; and, an appropriate spacing of stimuli corresponds with equal differences along the rating scale. The technique has been formally outlined by Stevens and Galanter (1957, pp. 381–382).

Figure 1 presents a hypothetical category estimation. The ordinate is the mean category rating. The abscissa is the stimulus scale between a lower and upper limit. This scale may represent a

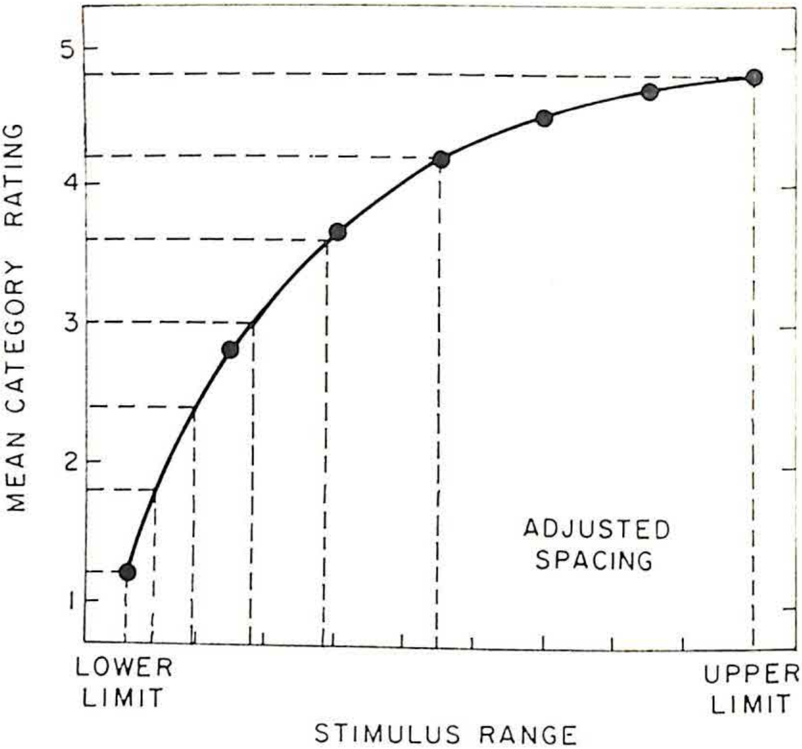


FIG. 1. Illustration of the application of the technique of adjusted spacing for a hypothetical category estimation. (See text for details.)

variable with equal interval or equal ratio properties, a variable with only ordinal properties, or a nominal set of stimuli that have been arranged only after the initial scaling to provide a smooth monotonic function. With the upper and lower limits of the range fixed, the following steps are carried out to obtain new stimuli between the limits:

1. Plot the initial category scaling.
2. Determine the difference between the mean category ratings assigned to the upper and lower stimulus limits.
3. Divide the difference by $(n - 1)$, where n is the number of stimuli desired for iterative scaling.
4. Add the quotient of Step 3 successively to the mean category rating assigned to the lower limit until the rating assigned to the upper limit is reached.
5. Draw a line parallel to the abscissa from each of the values of Step 4 to the smoothed curve.
6. Drop a perpendicular from the intersection to the abscissa.
7. Read the stimulus values on the abscissa scale.

Technique of average rating.—This technique attempts to achieve quickly the same end as the technique of adjusted spacing, when two or more sets of stimuli have been employed with common limits. The technique assumes that the average of two category ratings is less deviant from an unbiased rating than the more deviant rating alone. The following steps are carried out:

1. Plot the initial category estimations upon a single graph.
2. Average the ratings at corresponding stimuli.
3. Obtain a new set of stimuli by the method of adjusted spacing.

Technique of unequal stimulus frequencies.—This technique attempts to achieve an unbiased function when additional stimuli cannot be selected. Instead of adjusting the spacing, the technique adjusts the frequency of occurrence

of a fixed set of stimuli. The technique assumes that an appropriate frequency of stimulus presentation corresponds with equal differences upon the rating scale. The following steps are carried out:

1. From the initial category scaling, note the scale rating for each stimulus in terms of a monotonic category function.
2. For any stimulus *s*, calculate the average difference between the ratings assigned to *s* and its adjacent stimuli, or to the adjacent stimulus for the upper and lower stimulus limits.
3. Weight the relative frequency of occurrence of each stimulus by the average differences in weightings. For example, if four successive stimuli are associated with mean ratings of 1, 2, 5, and 7, the four stimuli would be presented with relative frequencies of 1, 2, 2.5, and 2, respectively.

Technique of equal-response frequencies.—This technique attempts to achieve the same end as the technique of adjusted spacing. The technique assumes that an appropriate spacing of stimuli corresponds with equal spacings along the cumulative-response frequency scale. The technique, and related techniques, are illustrated in Pollack (1964).

Technique of single-stimulus presentation.—This technique logically avoids the problem of stimulus bias by presenting only a single stimulus to each *S* or by employing only the first response of each *S* in an extended series of judgments. It is an empirical question whether the technique yields the same empirical function as the other techniques. This technique has been widely employed in loudness scaling (Garner, 1958; Stevens & Poulton, 1956).

Technique of category production.—This technique attempts to yield a "reasonable" distribution of stimuli quickly. The *S* is instructed to produce, or select, a given number of stimuli, *s*, in addition to the two end stimuli, such that successive stimuli appear to be "equally spaced." If $(s + 2) = n$, the number of desired stimuli, no further step is necessary. If $(s + 2) \neq n$, interpolation of the stimulus scale is required. The method was suggested by G. Stone (Stevens & Galanter, 1957, pp. 392–393).

Method

Sets of grays examined.—Table 1 presents the percentage reflectance of the population of 29 grays from which sets of 11 grays were selected. The reflectances are only approximate because of differences in reflectance of different samples of the same gray paper after repeated use and handling. The qualified specification of the stimulus sets is not regarded as serious since only the rank ordering of the stimulus set, rather than the reflectances, was used. Moreover, by concentrating upon the order property of the sets of grays, we were able to explore the applicability of the techniques to "weak" stimulus metrics on ordinal scales.

Table 1 also presents the specific 11 grays employed as the initial sets in the various tests. For example, Sets I, J, K each represent a wide range of grays with intermediate overlapping grays. Sets L, M, N each represent a wide range of grays without intermediate overlap. Sets O, P each represent a medium range of grays. Sets Q, R, S each represent subpopulations in which some stimuli are unavailable (u in Table 1). And, Sets T–X each represent narrow ranges of grays.

Each set of grays may be represented graphically by a double rank ordering, as shown in Fig. 2F. The abscissa is the rank-order darkness from 1 (white) to 29 (black); the ordinate represents successive stimuli from Stimulus 1 to Stimulus 11. In Fig. 2F, for example, the open circles represent a set of predominantly light grays with a scattering of dark grays; the filled squares represent a set of predominantly dark grays with a scattering of lighter grays; and, the half-filled triangles represent a set of predominantly intermediate grays with a scattering of light and dark grays.

Experimental methods.—In the method of category rating or *category estimation* (S. S. Stevens, 1958, Table 2), *S* was instructed to assign a number from 1 to 7 to grays presented one at a time. Paper 1 (white) was shown as an example of the lower end of Category 1; and Paper 29 (black) was shown as an example of the upper end of Category 7. The *S* was told that equal brightness differences were to be reflected in equal numerical differences. The *S* was also told that some of the ratings might be used more often than others, that some might not be

TABLE 1

SUBSETS OF 11 GRAYS FROM AVAILABLE POPULATION OF 29 GRAYS

Paper No.	% Reflect.	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	72	*	*	*	*	*	*			*	*	*	*				
2	66	*			*							u	*				
3	64	*			*					*	*		*				
4	58	*			*							u	*				
5	58	*			*					*	*	*	*				
6	47	*	*	*	*			*	*			u	*	*			
7	43	*			*			*		*	*		*	*			
9	41				*			*			u	u		*			
9	36				*			*	*	*		*		*			
10	32				*			*			u	u		*			
11	31					*		*		*	*			*			
12	30	*	*	*		*		*	*		u	u		*	*		
13	30		*			*				*		*			*		
14	24		*			*				u	u	u			*		
15	21		*			*		*	*		*				*		
16	20		*			*				u	u	u			*		
17	14		*			*				*		*			*		
18	13	*	*	*		*		*	*	u	u				*	*	
19	13					*			*		*	*				*	
20	12						*		*	u	u					*	
21	10						*	*	*	*		*				*	
22	9.2						*		*	u	u					*	
23	7.4			*			*		*		*	*				*	*
24	5.7	*	*	*			*	*	*	u						*	*
25	5.7			*			*			*	*	*					*
26	5.4			*			*			u							*
27	5.2			*			*				*	*					*
28	4.2			*			*			u							*
29	3.5	*	*	*	*	*	*			*	*	*					*

Note.—u = unavailable in the subpopulation.

used, and that the initial white and black papers would be included within the stimulus series.

In the method of *category sorting*, a set of 11 grays was successively arranged from white to black, or from black to white. The *S* was instructed to place the grays into seven, or fewer, piles so that successive piles represent approximately equal differences in brightness.

In the method of *magnitude selection*, a set of 11 grays was haphazardly arranged upon a white background. The *S* was instructed to select that paper which was most nearly half as bright (or twice as dark) as the white background. In other tests, the grays were arranged upon a black background and *S* was instructed to select that paper which was most nearly twice as bright (or half as dark) as the black background. Since there were no consistent differences between judgments under brightness or darkness instructions, results were pooled over instructions.

In the method of *ideal selection* (Coombs, 1958), a set of 11 papers was haphazardly arranged upon a white background. The *S* was instructed to select that paper which most nearly corresponds to "ideal" gray.

Experimental details.—Table 2 summarizes the experimental details for the separate experiments. The first column lists the graphical location of the results, with special reference to the double-rank representation of the stimuli employed. The second column

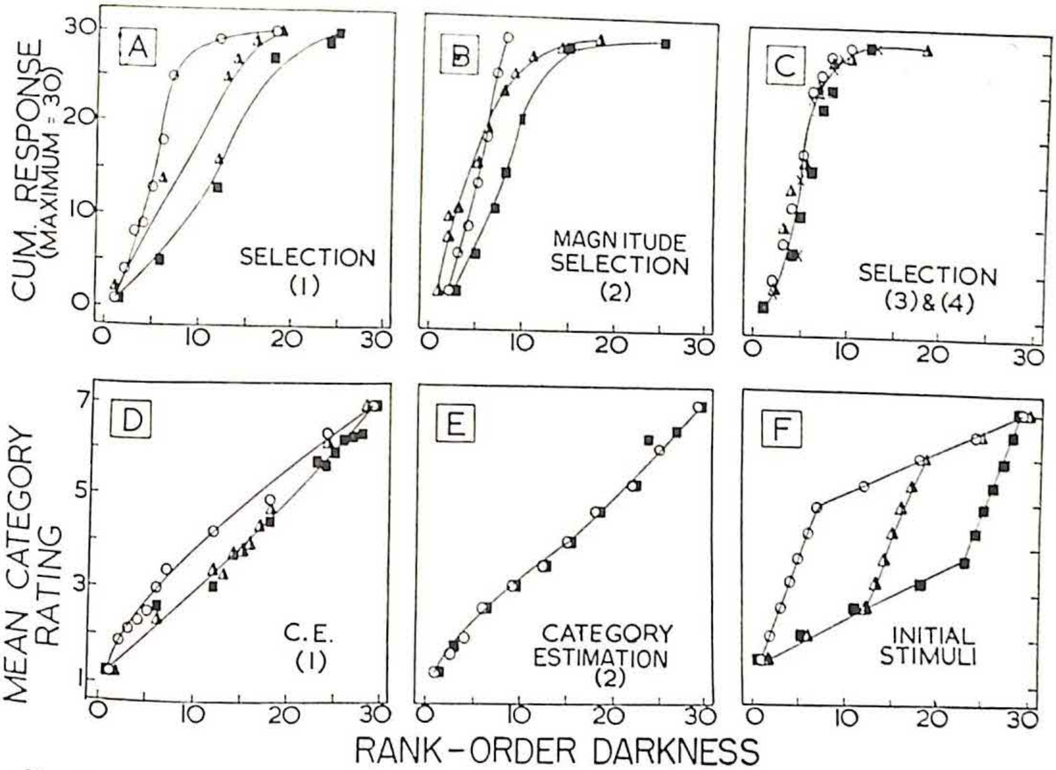


FIG. 2. Application of the technique of equal-response frequencies to magnitude selection (Fig. 2A, B, and C); and application of the technique of adjusted spacing to category estimation (Fig. 2D and E) for a wide range of gray papers (Sets I, J, K). (The double rank-order representation of the three initial sets of stimuli is presented in Fig. 2F.)

lists the initial sets of stimuli upon the first presentation, as coded in Table 1. The third column lists the number of initial sets of stimuli to which *S* was exposed. The fourth column lists the number of *S*s represented by each datum point, and whether *S*s were

TABLE 2

SUMMARIZATION OF EXPERIMENTAL DETAILS

Fig.	Sets	No. of Sets	*S*s	Iter./S	Exper. Proc.	Iter. Tech.
2F	I, J, K	1	30 U	1	MS CE	ER AS
3F	L, M, N	2	11 R 12 R	3 3	CS CE	AS AS
4	O, P	m	10–17 R	m	CE	AS
5E	Q, R, S	m	14–20 R	m	CE	AS
6	L, M, N	m	12–18 R	m	CE	US
7	I, J, K	m	12–18 R	m	CS CP CE	AS AS
8	all 29	—	14 R	1	CP	CP
9	—	—	25 U	1	CE	SS
10	T–X	1	6–7 R	5	MS	AS
—	I, J, K	1	15 U	1	IS	ER

Note.—m = mixed for different *S*s.

British Naval Ratings (R) or Cambridge University students (U). The fifth column lists the number of successive iterations to which each *S* was exposed, where 1 indicates that separate *S*s were employed for each iteration. The sixth column lists the experimental procedure: magnitude selection (MS), category estimation (CE), category selection (CS), category production (CP), or ideal selection (IS). The seventh column lists the iterative technique employed: equal-response frequencies (ER), adjusted spacing (AS), unequal stimulus frequencies (US), category production (CP), and single-stimulus presentations (SS).

All testing was performed individually. The number of rating categories in the category-estimation and sorting tests was 7; only a single selection was made in the magnitude and ideal selection tests. When successive iterations were carried out by the same *S*, revised stimuli were based upon the individual *S*'s responses. Within each test series, the illuminating conditions and background reflectance were constant. For ideal selection tests: the illuminating condition was outdoor, daylight, and the background reflectance was 68%. Other tests with Cambridge University students employed night room light and diffuse lamp light. Tests with British Naval Ratings employed spot illumination with the bulb about 15 in. over the test surface. Magnitude-selection tests employed a white background (Paper No. 1) or a black background (Paper No. 29). Category-estimation and category-selection tests employed a background reflectance of 68%.

Experimental strategy.—The basic strategy of experimentation was to deliberately introduce stimulus bias, by employing two or more weighted sets of stimuli; and then to attempt to remove, or neutralize, the stimulus bias by application of an iterative procedure. The application was considered successful when the same rating scale was obtained, despite large differences among the initial sets of stimuli. For example, three sets of grays with no intermediate overlap are represented in Fig. 3F. The three sets of grays led to substantially different initial category estimations, plotted as CE (1) in Fig. 3D where the three shapes reflect the three sets of stimuli of Fig. 3F. Application of the iterative technique of adjusted spacing upon the results of Fig. 3D led to three new sets of stimuli. The category estimations associated with the new sets are presented in Fig. 3E as CE (2). A single category estimation adequately describes the results of all sets. A further iteration, CE (3) in the form Xs, demonstrates that the cate-

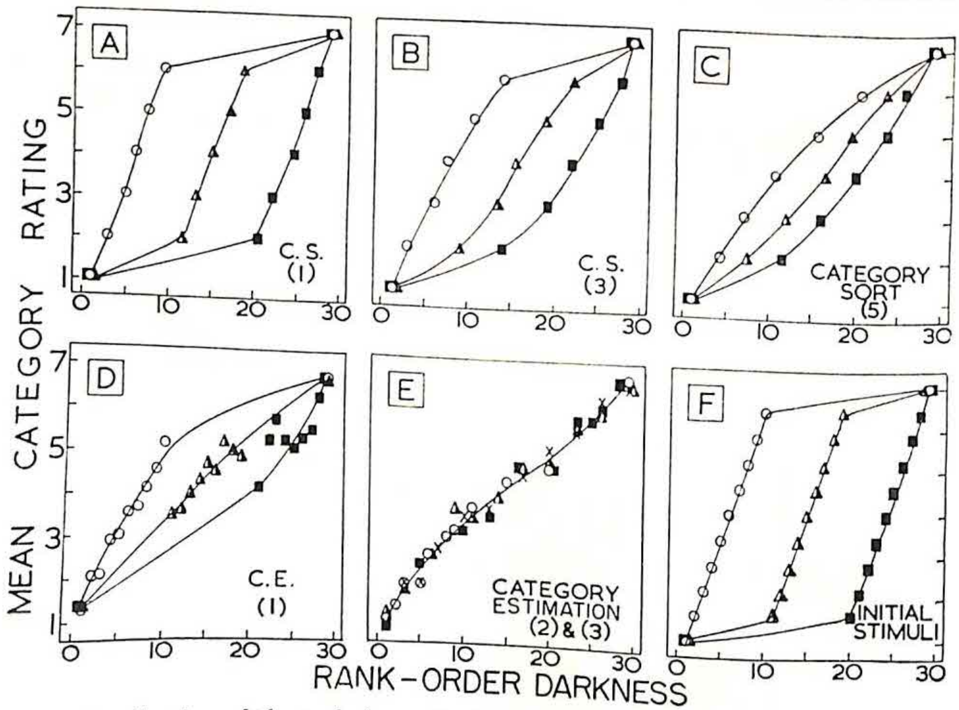


Fig. 3. Application of the technique of adjusted spacing to category sorting (Fig. 3A, B, and C) and to category estimation (Fig. 3D and E) for the results of individual *S*s for a wide range of gray papers (Sets L, M, N). (Initial stimulus sets in Fig. 3F.)

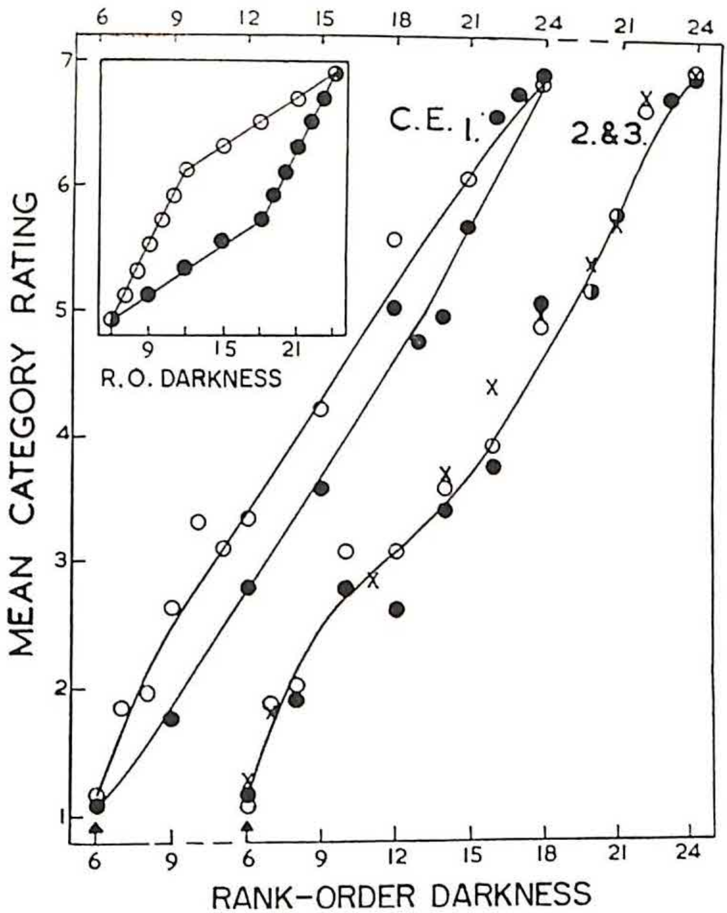


FIG. 4. Application of the technique of adjusted spacing for the category estimation of a restricted range of gray papers.

gory estimation is stable upon further iteration. This totality of operations will be described as the *convergence* upon a stable category estimation. By contrast, under the procedures illustrated in Fig. 3A, B, C, successful convergence was not achieved within five successive iterations.

RESULTS

The main findings may be briefly summarized: convergence upon a stable empirical function is quickly obtained by all techniques under almost all of the conditions investigated. Convergence may be substantially slower where the following restrictions are simultaneously applied: a highly weighted set of stimuli (Sets L, M, N) or narrow set of stimuli (Sets T through X); an experimental method which implicitly or explicitly imposes strong restraints upon *S* (sorting or selection); and, the application of the iterative technique to the individual *S*'s results, rather than application to the average *S*'s result. These conditions apply to the results of Fig. 3A–C and Fig. 10.

The overview may be made more complete by the following additional details:

Sample results for additional variables are represented in Fig. 6 and 8. The initial category estimations for these variables were similar to those of Fig. 2A and 3D.

In Fig. 2C, 3E, 4, and 7, Xs represent the results of an additional iteration which tested the stability of

the resulting empirical function. In additional tests, rapid convergence of the category estimations under the conditions of Fig. 2 was obtained whether separate groups were employed for each iteration of each stimulus set, or whether a single group of *S*s was followed over successive iterations. In additional tests, the slow rate of convergence of the category-sorting tests in Fig. 3A, B, and C was hastened by simply applying the technique of adjusted spacing to the pooled sortings across *S*s, rather than to the individual data alone.

The results of Fig. 4 suggest that well-defined stimulus end points, such as black and white, are not necessary for rapid convergence upon a stable category estimation.

The rapid convergence of the category estimations for each of the initial subpopulations of 21 papers of Fig. 5A, B, C is not unexpected. More important when the third category estimations are plotted against the common scale of 29 papers, a common category estimation is achieved across subpopulations (Fig. 5D). The results suggest that the category estimation associated with a given stimulus, resulting from iterative techniques, will not be critically determined by the arbitrarily defined population from which the set of stimuli is drawn. The population must only provide stimuli for interpolation to the technique of adjusted spacing in regions of high discriminability.

Unlike the other inserts which represent initial stimulus sets, the insert of Fig. 6 represents the distribu-

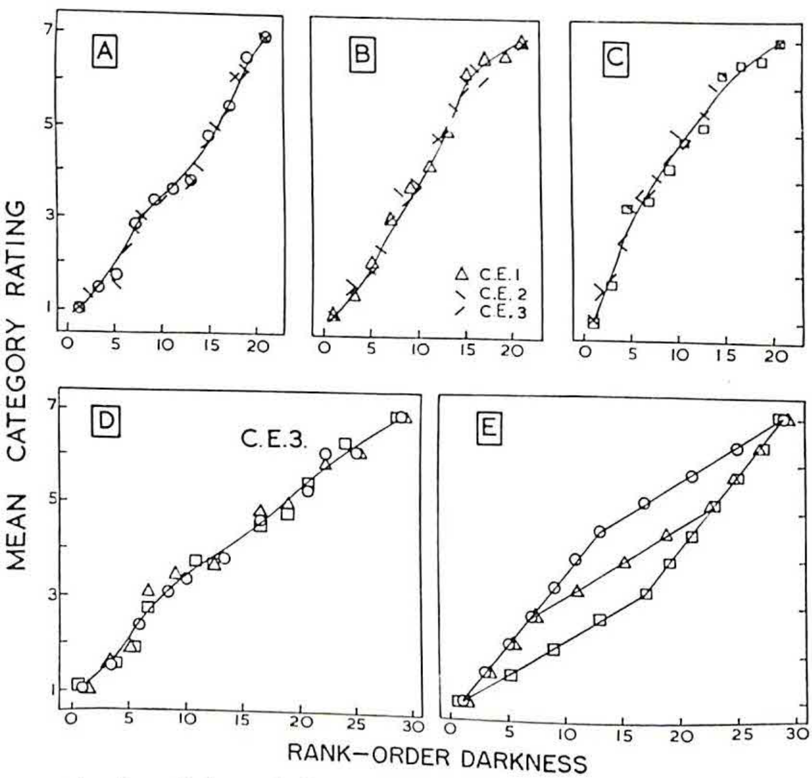


FIG. 5. Application of the technique of adjusted spacing for the category estimation of gray papers within different initial subpopulations. (The initial category estimations of the distributions of Fig. 5E are presented in Fig. 5A, B, and C in terms of each subpopulation. The third category estimation of Fig. 5A, B, and C is plotted in Fig. 5D against the common scale of 29 papers.)

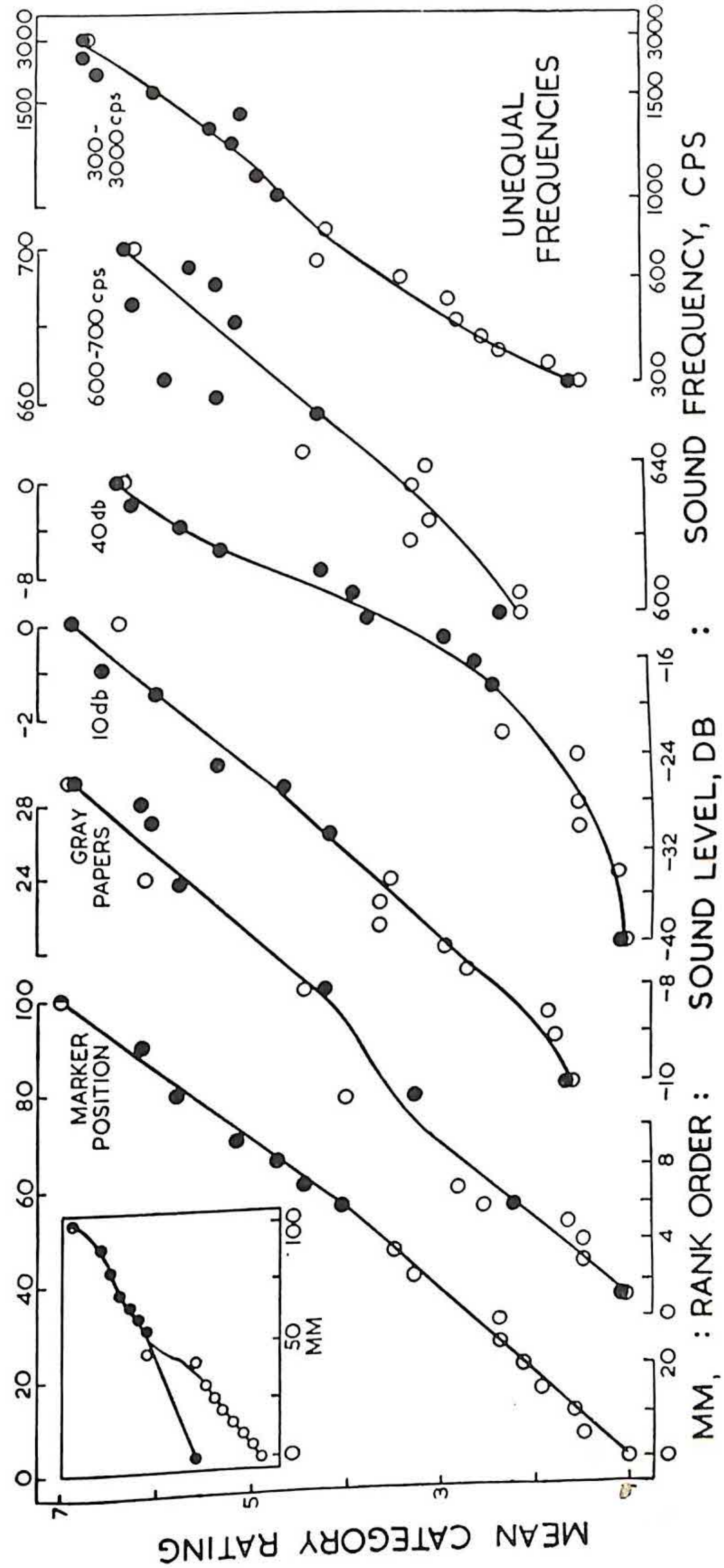


FIG. 6. Application of the technique of unequal stimulus frequencies to the category estimation of four variables.

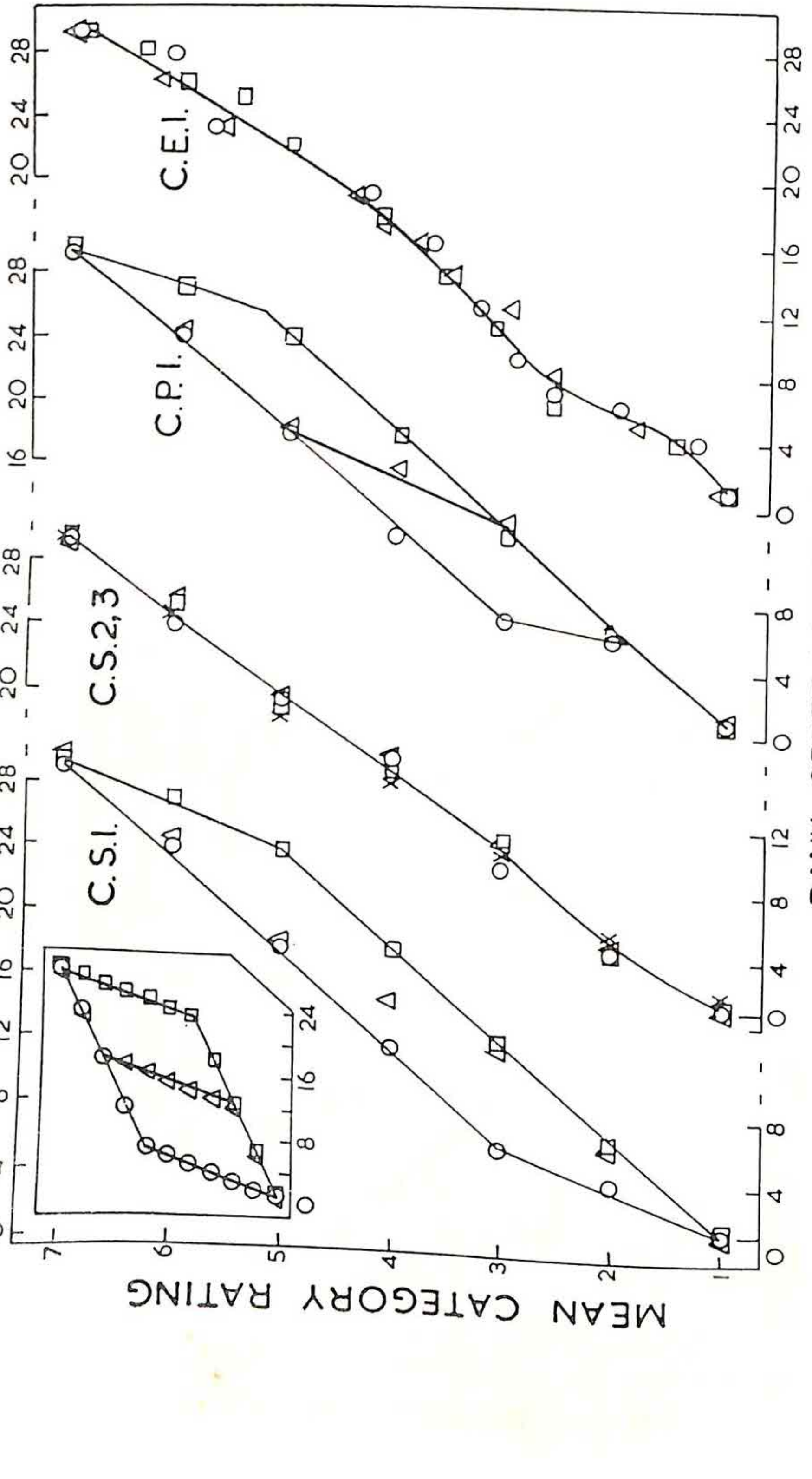


FIG. 7. Application of the techniques of adjusted spacing and category production to the category sorting and category estimation of gray papers.

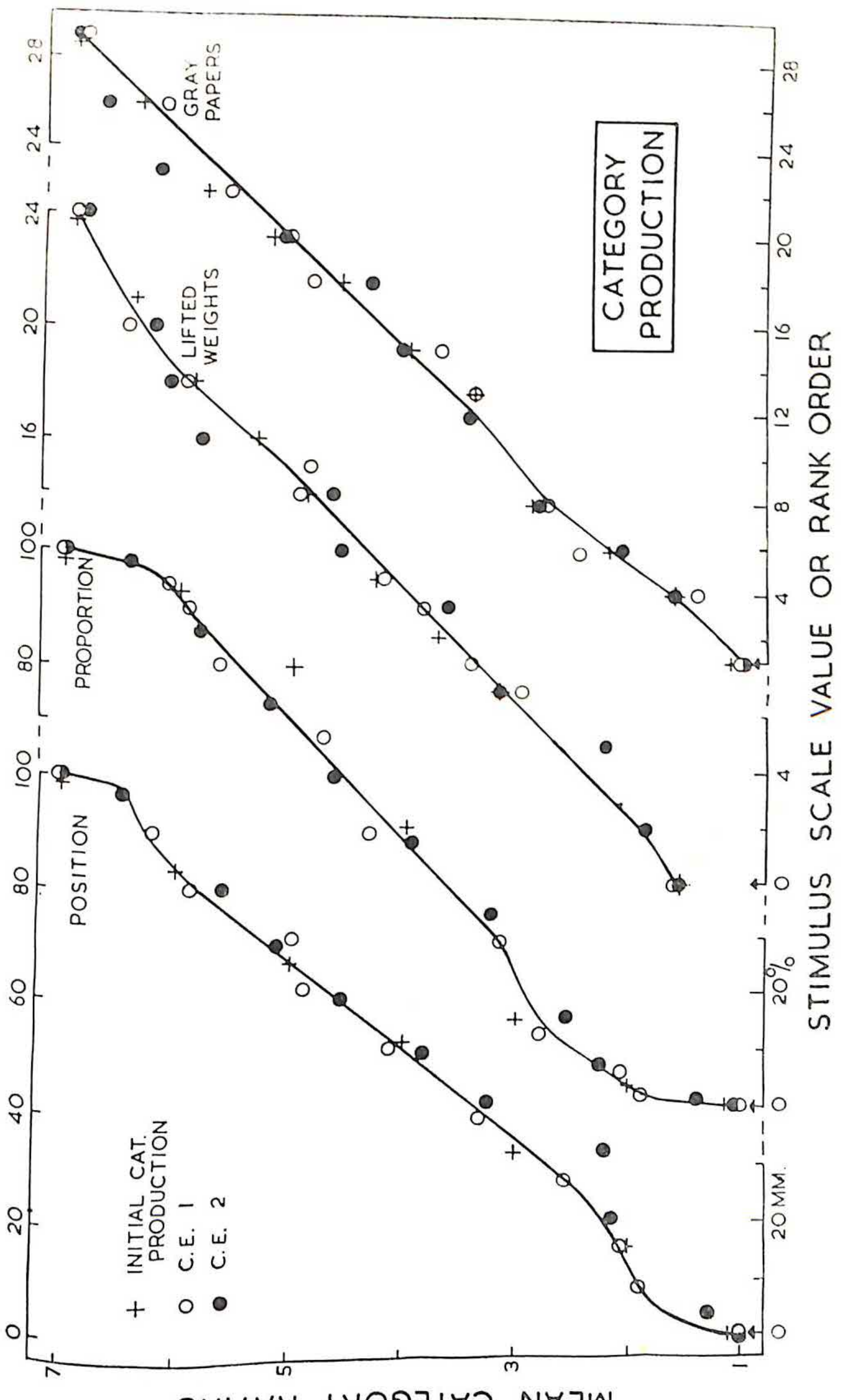


Fig. 8. Application of the technique of category production to the category estimation of four variables.

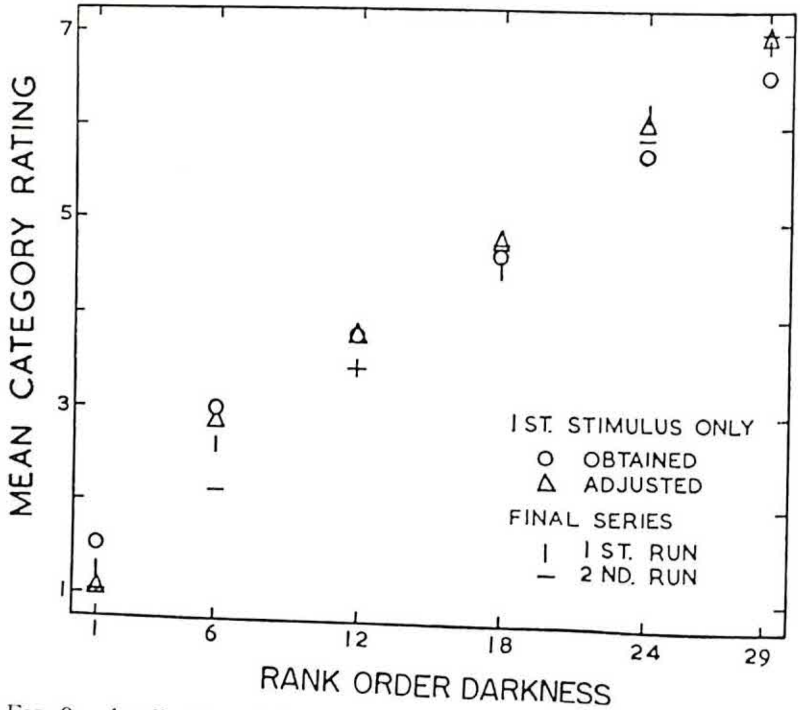


FIG. 9. Application of the technique of single-stimulus presentation to the category estimation of gray papers.

tion of unequal stimulus frequencies following the initial category estimations. In the case illustrated, two nonoverlapping sets of marker positions were employed.

The tests represented in Fig. 7 suggest that rapid convergence upon a stable category sorting (CS) is possible if the initial stimulus sets are partially overlapping, rather than nonoverlapping as in Fig. 3. Category production (CP) from among the initially weighted Sets I, J, K yielded sets of stimuli which led directly to a stable category estimation (CE), based upon the previous category productions.

Category productions that were obtained with examples from the entire stimulus range led directly to stable category estimations for a wide range of variables (Fig. 8).

The ratings obtained with the technique of single-stimulus presentation (Fig. 9) are in fair agreement with the mean category estimations after several iterations with a set of 11 grays. There was an apparent tendency for *S* to withhold extreme ratings to the initial stimulus presentation. If the initial ratings are adjusted to the extremes of the rating scale, closer agreement is obtained between the two techniques. Comparison between the techniques is made difficult by differences in category estimations furnished by the first and second runs through the set of 11 grays.

Convergence is slow when *S* must make selections from a narrow stimulus range (Fig. 10). The slow rate of convergence must be viewed in light of the severe restrictions imposed by the narrow stimulus range. A portion of the range of grays (indicated by the stippled area of Fig. 10) was unattainable, even after five successive extreme selections. For example, if the set of seven grays were centered on Paper 26 for the first selection, Paper 10 was

still not attainable after five successive selections. The reference line in Fig. 10 indicates the result expected had the initial sets completely determined the final result.

The median selections for "ideal" gray were Papers 11, 12, and 12.5 for Sets I, J, K, respectively. Presentation of a new set of stimuli, obtained by the technique of equal-response frequencies from the initial ideal selections to Set I, yielded a median selection at Paper 12. The results suggest that convergence upon a stable ideal selection may be quickly achieved by iterative procedures.

Discussion

At first, the several iterative techniques appear to act in a magical way to fashion stable rating scales. There is nothing magical about their action. In the technique of adjusted spacing, for example, new stimuli are obtained on successive iterations. The rating scales become more and more alike on successive iterations because the stimuli become more and more alike. Perhaps the only requirement for convergence upon a stable rating scale is that *S*'s responses reflect something more than the rank-order properties of the stimulus set alone.

The failure of the category-sorting procedure with a wide range of nonover-

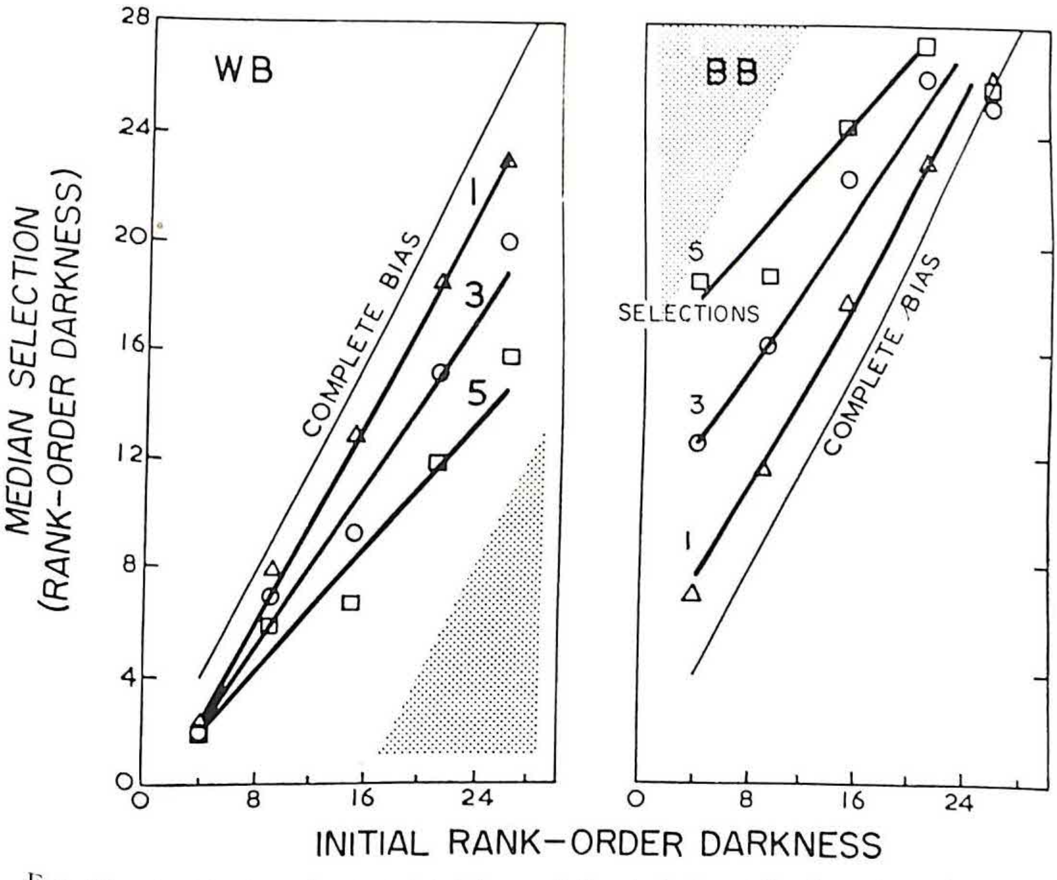


FIG. 10. Application of a restricted form of the technique of adjusted spacing to the magnitude selection of grays. (Tests upon the white background [Paper No. 1] are presented on the left; the black background [Paper No. 29] on the right. Seven successive papers were initially available for selection, centered upon Papers 4, 9, 15, 21, and 26, respectively. The parameter is the number of successive selections. The shaded areas represent the range of unattainable conditions even after five successive selections of the extreme gray of each set of seven. Within each vertical line of points, the median of the individual selections is presented.)

lapping grays probably relates to the rank-order constraints imposed by the category-sorting procedure. In order to investigate this point, additional tests were carried out with sets of 11 cards marked with the numerals 1, 3 . . . 29, according to Sets L, M, and N. The *Ss* (Royal Naval Ratings) were instructed to sort the cards into seven, or fewer, piles so that the "average numerical distance from pile to pile was approximately constant." The instructions were apparently disregarded because large gaps in the numerical ratings were not reflected in the sortings. A typical result for sorting the numerals of Set L into successive piles was: 1 & 2; 3; 4 & 5; 6; 7 & 8; 9 & 10; 29. That is, *S* interpreted the instructions in such a way as to produce seven, and no fewer, piles whether dealing with grays or numerals. The agreement between the rank-order brightness and the rank-order numerical designation was high. Over 75% of the sorting categories agreed exactly for Sets L, M, N and over 65% of the sorting categories agreed exactly for Sets I, J, K. In less than 2% of the cases did numerical and brightness-category ratings differ by more than one category rating. This result suggests that the category-sorting method closely reflects the rank ordering of the stimuli. It may be noted that even with the strong restrictions of the category-sorting procedure, convergence upon a stable rating function may be achieved if the initial sets of stimuli are overlapping (CS tests of Fig. 7), or if successive iterations are based upon the pooled responses of several *Ss* rather than upon the responses of individual *Ss*.

The minor difficulties with the exceptional cases, however, should not lead us from the main result: rapid convergence upon an unbiased scale is achieved by a wide range of iterative techniques, for a wide range of stimulus variables and for a wide range of psychophysical methods.

REFERENCES

Arons, L., & Irwin, F. W. Equal weights and psychophysical judgments. *J. exp. Psychol.*, 1932, **15**, 733–751.

Blough, D. A. Method for tracing dark adaptation in the pigeon. *Science*, 1955, **121**, 703–704.

Coombs, C. On the use of inconsistency of preferences in psychological measurements. *J. exp. Psychol.*, 1958, **55**, 1–7.

Garner, W. R. Context effects and the validity of loudness scales. *J. exp. Psychol.*, 1954, **48**, 218–224.

Garner, W. R. Half-loudness judgments without prior stimulus context. *J. exp. Psychol.*, 1958, **55**, 482–485.

Helson, H. *Adaptation level theory.* New York: Harper & Row, 1964.

Hirsh, I. J. Békésy's audiometer. *J. Acoust. Soc. Amer.*, 1962, **34**, 1333–1336.

Johnson, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

Parducci, A. Range-frequency compromise in absolute judgment. *Psychol. Monogr.*, 1963, **77**(2, Whole No. 565).

Pollack, I. Neutralization of stimulus bias in auditory rating scales. *J. Acoust. Soc. Amer.*, 1964, **36**, 1272–1276.

Poulton, E. C., & Stevens, S. S. On the halving and doubling of the loudness of white noise. *J. Acoust. Soc. Amer.*, 1955, **27**, 329–331.

Rosenthal, R. On the social psychology of the psychological experiment. *Amer. Scientist*, 1963, **51**, 268–283.

Smith, K. E. Stimulus programming *in* psychophysics. *Psychometrica*, 1961, **26**, 27–33.

Stevens, J. C. Stimulus spacing and the judgment of loudness. *J. exp. Psychol.*, 1958, **56**, 246–250.

Stevens, S. S. On the averaging of data. *Science*, 1955, **121**, 113–116.

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196.

Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

Stevens, S. S., & Poulton, E. C. The estimation of loudness by unpracticed observers. *J. exp. Psychol.*, 1956, **51**, 71–78.

Stevens, S. S., Rogers, M. S., & Herrnstein, R. T. The apparent reduction in loudness: A repeat experiment. *J. Acoust. Soc. Amer.*, 1955, **27**, 326–328.

(Received March 9, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 579–582

REWARD VERSUS NONREWARD IN A SIMULTANEOUS DISCRIMINATION [1]

R. ALLEN GARDNER [2] AND W. B. COATE [3]

Wellesley College

The procedure of a 2-choice discrimination was varied in 2 ways. Under M+ conditions, 60 rats were trained on problems in which S+ varied from trial to trial while the same S− was presented on every trial. Under M− conditions, 60 additional rats were trained on problems in which S− varied from trial to trial while the same S+ was presented on every trial. Performance was poorer under the M+ condition than under the M− condition, indicating that reward for choices of a specific S+ was a more critical determinant of discriminative performance than was nonreward for choices of a specific S−.

In a standard procedure of simultaneous-discrimination training the same two discriminanda are presented on every trial, and choices of S+ are rewarded while choices of S− are not rewarded. Discrimination learning is measured by increases in R+, the probability of choosing S+, or, symmetrically, by decreases in R−, the probability of choosing S−. With only two discriminanda and two response categories, however, the relative effects of reward for choices of S+ and nonreward for choices of S− cannot be evaluated because increases in R+ are completely confounded with decreases in R−.

In the present experiment, two variants of the standard procedure were used. One variant was a Multiple-Positive procedure (M+) in which different positive discriminanda were presented on different trials, but the same S− was presented on every trial. The second variant was a Multiple-Negative procedure (M−) in which different negative discriminanda were presented on different trials, but the same S+ was presented on every trial. Both M+ and M− should be more difficult than the standard procedure in which the same S+ and the same S− are presented on every trial. But, of the two multiple procedures, which should be the more difficult? In M+, additional difficulty is introduced via the conditions of reward training; in M−, via the conditions of nonreward training. We reasoned that the relative difficulty of M+ and M− should reflect the relative importance of reward training and nonreward training, if the two multiple procedures could be equated in all other respects. Accordingly, the following experiment was designed to compare the M+ and M− procedures under a specific set of experimental conditions.

METHOD

Subjects.—The *S*s were 120 male albino rats of the Sprague-Dawley strain, 90–120 days old, purchased from the Charles River Breeding Laboratories.

Apparatus.—The experimental compartment used for Discrimination Training is shown in Fig. 1. It consisted of a choice chamber and a treadle chamber connected by a narrow passageway and covered by a lid. The choice chamber was divided in half by a partition which was attached to the front wall. Two 2.25-in. circular openings (P) were cut out of the front wall of the choice chamber

[1] This research was supported by Grant M-4337 from the National Institute of Mental Health, United States Public Health Service.

[2] Now at the University of Nevada.

[3] Now at Hazleton Laboratories, Falls Church, Virginia.

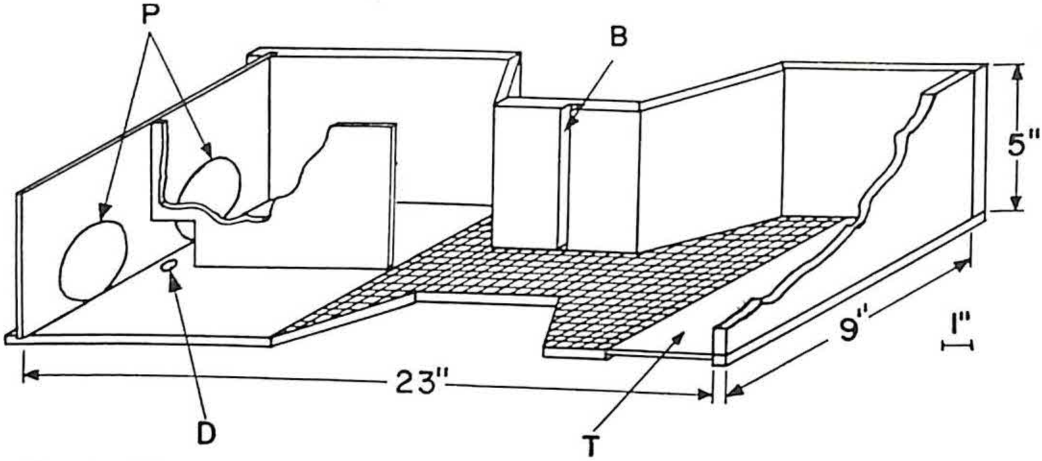


FIG. 1. Diagram of the experimental compartment used for Discrimination Training.

on either side of the partition. Each opening was covered by a translucent Plexiglas panel, hinged at the top so that it could be pushed outwards to actuate a microswitch. The two panels also served as screens on which the discriminanda were projected from behind by an automatic filmstrip projector (not shown). When the panels were dark, both could be lighted by momentarily depressing the treadle (T), and both were darkened again by the first push on either panel. A correct choice was rewarded with a bead of water, delivered automatically by a solenoid operated dipper through an aperture (D) in the floor of the choice chamber. Actuation of the dipper solenoid was accompanied by a flash of light on the correct panel, and a cutout in the partition permitted *S* to take rewards from either side.

A second compartment, used for Preliminary Training, had the same floor plan as the compartment shown in Fig. 1. The front wall of the Preliminary Training compartment could be removed and replaced by a front wall with only one circular opening and panel in its center. The panels used in Preliminary Training were lighted by lamps from outside the compartment, and when lighted, these panels presented a homogeneous white field.

The experimental rooms were maintained in darkness with a sound screen of white noise, and the automatic control and recording apparatus were kept in a separate room.

Deprivation.—Dry food was available in the living cages at all times. For 6 days before the beginning of Preliminary Training each *S* was placed on a 23-hr. schedule of water deprivation. During Preliminary Training and Discrimination Training each *S* was given free access to water in his living cage for 45 min. each day, and was 22.5 hr. deprived of water at the start of each daily training session.

Preliminary Training.—Each *S* was handled once daily on each of 4 days before the beginning of Preliminary Training. Preliminary Training consisted of five, 35-min., daily sessions of training which were designed to approximate, by stages, the procedure of Discrimination Training. During Days 1–3, the single-paneled front wall of the choice chamber was in place. On Day 1, *S* was confined to the choice chamber by a block placed at B in Fig. 1. The single panel remained lighted throughout the session and all panel pushes were rewarded. On Days 2 and 3, the block was removed and *S* was free to move through the passageway between the choice chamber and the treadle chamber. When the panel was pushed while it was lighted, a reward was delivered and the panel was immediately darkened. The panel was relighted when *S* left the choice chamber, entered the treadle chamber, and depressed the treadle. On Days 4 and 5, the single-paneled front wall was removed and replaced by the double-paneled wall with its partition as shown in Fig. 1. Now, when *S* depressed the treadle, one panel was lighted and the other remained dark. Pushing the lighted panel was rewarded, pushing the dark panel was not rewarded, and pushing either panel darkened both.

Discrimination Training.—There were 432 trials of Discrimination Training, 48 on the first day and 96 on each of 4 succeeding days. The discriminanda were alternating black and white stripes, two black and three white, and each stripe was $\frac{9}{16}$ in. wide at the panel. The discriminanda differed only in the angular

orientation of the stripes, and four angles were used: 0°, 45°, 90°, and 135°. One of the angles (Key Angle) was projected on one panel, and one of the three remaining angles (Alternative Angle) was projected on the opposite panel during each trial. Each Alternative Angle was presented on one third of the trials, the sequence of presentation and left/right placement of the Alternative Angles being counterbalanced and evenly distributed over each block of 48 trials.

Experimental design.—Half of the 120 *Ss* were trained under the Multiple-Positive (M+) condition, in which all choices of the Alternative Angles were rewarded and no choices of the Key Angle were rewarded. The other half of the *Ss* were trained under the Multiple-Negative (M−) condition, in which all choices of the Key Angle were rewarded and no choices of the Alternative Angles were rewarded. Each of the four angles was used as the Key Angle for an equal number of *Ss* in both conditions. Fifteen *Ss* were randomly assigned to each of the resulting eight subgroups in such a way that the time of day and week of the experiment during which they were trained were roughly equated.

Thus, each M+ animal had, in effect, three discrimination problems to solve concurrently, each with a different S+ but all three with the same S−; while, subgroup for subgroup, each M− animal had three equivalent problems, all with the same S+ but each with a different S−.

Results

During the last 48 trials of Discrimination Training (Trials 385–432) the median number of errors was 10.0 for the M+ group and 3.5 for the M− group. By the median test (Siegel, 1956, pp. 111–116) this difference yielded χ^2 (1) = 5.63, $p < .025$.[4] Performance also depended upon which angle served as the Key Angle, χ^2 (3) = 33.1, $p < .01$; the discrimination problems were more difficult when the Key Angle was 45° or 135° than when it was 0° or 90°. There was less difference between the M+ and M− conditions when the Key Angle was 45° or 135°, but this was confounded by the fact that poorer performance during the last 48 trials was also closer to chance and hence less reliable.

Although training was terminated after 432 trials, the results were also tabulated in terms of trials to criterion. For each *S*, the number of trials required to make a given number of correct choices out of any 48 consecutive trials was tabulated for each criterion from 25/48 to the best performance reached within the allotted 432 trials. The median number of trials required for each criterion from 25/48 to 45/48[5] was greater for the M+ group than for the M− group, and this difference was significant at the .05 level or better for each criterion from 38/48 to 44/48 by the median test. Since the criterion of 44/48 was reached by half of the *Ss* in the total sample, we can say that the difference would have been at least as significant if all 120 *Ss* had been run to a criterion of 44/48.

Discussion

Assuming that reward for each choice of a particular S+ increases the probability of choosing that S+ more than it increases the probability of choosing any other stimulus object, the optimal stimulus condition for reward training would be one in which precisely the same S+ were presented on every trial so that all rewards would be received for choices of that S+. Thus, the effectiveness of the rewarded trials was reduced by the M+ procedure because different positive discriminanda were presented on different trials. Similarly, the effectiveness of the nonrewarded trials was reduced by the M− procedure because different negative discriminanda were presented on different trials. In all other respects, the M+ and M− discrimination problems were equated.

[4] Analysis of variance performed on these data yielded an F (1, 112) of 6.11, $p < .025$.

[5] 45/48 was the highest criterion that was reached or exceeded by half of the M− group.

Under the conditions of the present experiment, poorer performance was found under the M+ procedure than under the M− procedure. This finding is consistent with models of discrimination learning which assign greater weight to the effects of reinforcement than to the effects of extinction (Estes, 1959; Spence, 1936); but it is not consistent with Harlow's (1959) proposition that discrimination learning can be described as a "uniprocess" consisting of inhibitory learning, or with Amsel's (1962) conclusion that "avoidance of nonreward is a more powerful factor in discrimination than approach to reward [p. 309]."

REFERENCES

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning. *Psychol. Rev.*, 1962, **69**, 308–328.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 392–469.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 526–533.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

(Received March 11, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 583–589

EFFECTS OF FOOD AND WATER DEPRIVATION ON THE PERFORMANCE OF A RESPONSE MOTIVATED BY ACQUIRED FEAR[1]

RONALD LEY[2]

Syracuse University

On the basis of Hull's drive-summation hypothesis, it was predicted that a fear-motivated response would be augmented by the addition of an irrelevant appetitional need or a combination of irrelevant needs and that the level of augmentation would be a direct function of the level of deprivation. 84 rats were fear conditioned, after which they were put on a 23-hr. food- and water-deprivation schedule. On the 7th and 8th days following this, Ss were given 25 trials of fear-motivated hurdle jumping under 6 different combinations of food and water deprivation. Median reciprocals of latency measures indicated that the rank order of performance on the 2nd day of testing and on the last block of 5 trials was consistent with Hull's drive-summation hypothesis.

The trends of the results of studies designed to measure the effects of irrelevant needs suggest that needs combine to increase motivation as measured by changes in performance (Amsel, 1950; Meryman, 1952). Such exceptions to this generalization as do occur seem to involve the combination of hunger and thirst (Bolles, 1961; Kendler, 1945; Verplanck & Hayes, 1953). Examination of these studies reveals that the response measure employed in each instance was either a consumatory response or an instrumental response learned on the basis of one of these needs. Since it seems reasonable to assume that hunger and thirst are not independent, it is conceivable that the failure to find a motivational effect is the result of its being masked by other factors such as drive-stimulus generalization decrements. Thus, if a procedure were used where the possibility of such decrements was minimized, it might be possible to measure the pure energizing effects of the combination of hunger and thirst if such exist. If appetitional needs combine with aversive needs as Amsel's and Meryman's findings suggest, it seems feasible to use the typical paradigm of acquired drive studies for this purpose. The present study attempted to measure the effects of food deprivation, water deprivation, and combinations of food and water deprivation on the performance of a response motivated by acquired fear. On the basis of Hull's (1943) drive-summation hypothesis it was predicted that a fear-motivated response would be augmented by the addition of an irrelevant appetitional need or combination of irrelevant needs and that the level of augmentation would be a direct function of the level of deprivation.

METHOD

Subjects.—The *S*s consisted of 84 naive, male, hooded rats from the colony maintained at the Psychology Department of Syracuse

[1] This paper is based upon a dissertation submitted to the Graduate School of Syracuse University in partial fulfillment of the requirements for the PhD degree. The author wishes to thank Wallace R. McAllister under whose direction this investigation was carried out.

This investigation was supported in part by a research grant, M-2064, from the National Institute of Mental Health, Public Health Service, to Wallace R. and Dorothy E. McAllister.

[2] Now at the Graduate Faculty, New School for Social Research.

University, 91–128 days of age on their first experimental day; average age was 108 days.

Apparatus.—Fear conditioning was performed in two separate white, wooden framed shock boxes, the interior dimensions of which were $9\frac{3}{4}$ in. long $\times$ $2\frac{1}{2}$ in. wide $\times$ 5 in. high. The floor of each box consisted of 21 $\frac{3}{32}$-in. brass welding rods spaced $\frac{7}{16}$ in. apart through which a 115-v. electric current (UCS) could be delivered at the rate of two impulses per grid per sec. A 100,000-ohm resistor was in series with each *S*. Hinged to the rear wall of each shock box was a 12-in. high box, the hardware cloth bottom of which served as the ceiling of the shock box. Mounted $\frac{3}{4}$ in. above the hardward cloth bottom was a pane of opal glass which provided diffuse transmission of light from the tunnel to the interior of the shock box. Concealed within the tunnel were the CS, a 40-w. incandescent lamp situated 8 in. above the opal glass, and the intertrial light, a $7\frac{1}{2}$-w. incandescent lamp situated 10 in. above the opal glass. The illumination provided by the CS was approximately 120 footcandles (ftc.) and the intertrial illumination, $7\frac{1}{2}$ ftc.

The hurdle-jumping apparatus consisted of two compartments separated by a $\frac{3}{4}$-in. partition containing a guillotine door $2\frac{1}{2}$ in. wide $\times$ 3 in. high which rested on a 2-in. high hurdle. The start box was built to be identical to the shock boxes, whereas, the safe box was gray and had a plywood floor. Raising of the manually operated guillotine door, closed a switch which resulted in the simultaneous activation of the CS and Hunter klockounter. A second switch was located beneath the safe-box floor. The weight of *S* on the floor closed the switch which simultaneously turned off the CS and stopped the klockounter.

Design.—The overall design of the study involved the fear conditioning of rats through the pairing of an increase in illumination (CS) with a 115-v. shock (UCS) delivered to *S*s by means of a grid floor. Immediately following 35 fear-conditioning trials, conducted while *S*s were maintained on an ad-lib. feeding schedule, all *S*s were put on a 23-hr. limited time food- and water-deprivation schedule. Seven days later each *S* was tested in the hurdle-jumping apparatus, 25 trials per day, for 2 days, under different levels of hunger and thirst. Response latencies for each *S* on each trial were converted to reciprocals, and mean reciprocals for each *S* were computed in blocks of 5 trials thus providing for five blocks of 5 trials the first day of testing and five blocks of 5 trials the second day of testing.

Procedure.—The 84 *S*s used in the experiment were introduced to the experimental setting in squads of 12, 2 *S*s of each squad randomly assigned to one of the six experimental groups. During the 11 days required for the administration of the experimental conditions, *S*s were maintained in individual living cages.

On Days 1 and 2 *S*s received $7\frac{1}{2}$ min. of handling and 20 min. of exploration of the hurdle-jump apparatus with intertrial light on; 10 min. in the start box and 10 min. in the safe box.

On Day 3 each pair of *S*s was administered 35 fear-conditioning trials. Following random assignment to the pair of shock boxes, the first trial began 1 min. after *S*s were placed in the boxes. Each trial began with the onset of the CS, 4 sec. after which the UCS was presented for a 2-sec. duration with CS and UCS terminating simultaneously. Intertrial interval was 2 min. Following the last fear-conditioning trial for the last pair of *S*s in each squad, food and water were removed from the home cages of each *S* in that squad. For the ensuing 5 days each *S* in the squad was maintained on a limited time feeding schedule with free access to food and water for a daily 1-hr. period.

On Day 9 (the sixth day of deprivation) the 1-hr. feeding period was adjusted so that it began exactly 24 hr. before the scheduled test trials of the next day. Since test trials were scheduled at 1-hr. intervals, the feeding schedule of Day 9 was similarly staggered at 1-hr. intervals coinciding for each group with their scheduled testing period. This resulted in slight variations from the regular feeding time ranging from 1 to 3 hr.

To provide for the designated combinations of levels of food and water deprivation at the onset of testing, the various treatment groups received food and water for 1-hr. periods on Day 10 in the following manner: Group C (23H-23T), food and water the hour following test trials; Group F (OH-OT), food and water the hour preceding test trials; Group A (23H-OT), water the hour preceding test trials and food the hour following test trials; Group E (OH-23T), food the hour preceding test trials and water the hour following test trials; Group B (23H-12T), water for a 1-hr. period 12 hr. before testing and food the hour following testing; Group D (12H-23T), food for a 1-hr. period 12 hr. before testing and water the hour following testing.

Each of the 25 test trials of Day 10 consisted of placing the first *S* of the pair in the start box of the hurdle-jumping apparatus, 10 sec. after which *E* raised the guillotine door

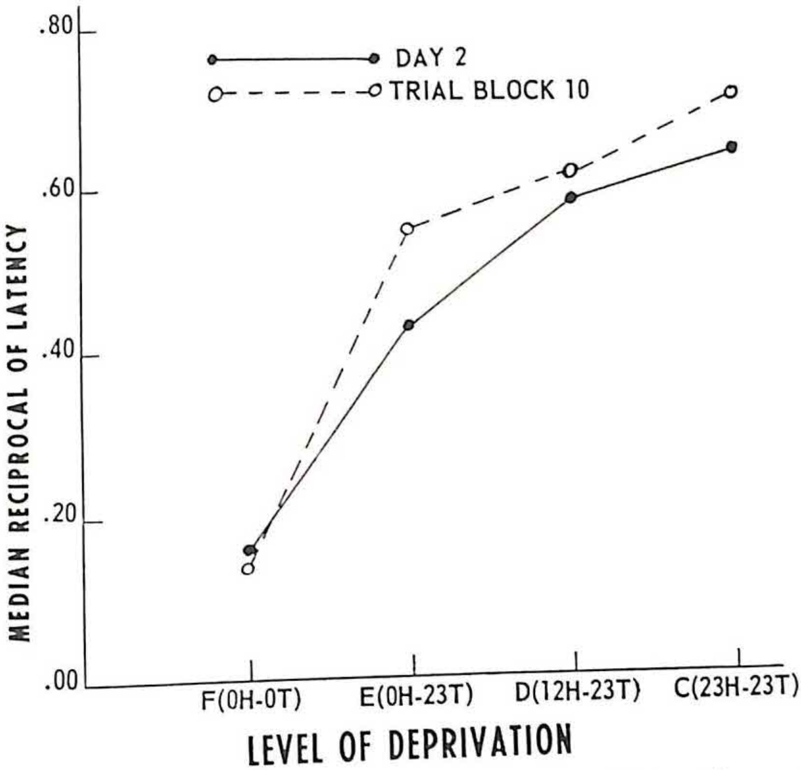


FIG. 1. Hurdle-jumping performance as a function of the addition of hunger to thirst.

which simultaneously activated the CS and klockounter. The UCS was not presented during test trials. When *S* jumped the hurdle, the depression of the floor switch of the safe box terminated the CS and stopped the klockounter, thus providing a measure of response latency. The guillotine door was lowered following *S*'s entrance into the safe box, where *S* remained for 10 sec., after which he was removed and placed in a holding cage while the second *S* of the pair was subjected to the hurdle-jump conditions. If *S* did not jump within 60 sec., *S* was removed from the start box, placed in the holding cage, and a latency of 60 sec. was recorded.

On Day 11, the final experimental day, each *S* was given 25 test trials identical in procedure to those administered on Day 10, except that the test period for each pair of *S*s was 1 hr. later than on the previous day.

RESULTS

The data of the experiment are presented as two separate but parallel studies. One study includes the data of four groups, three of which were tested under 23 hr. of water deprivation but under different levels of food deprivation; Groups E (0H-23T), D (12H-23T), and C (23H-23T); and one of which was tested under 0 hr. of food and water deprivation, Group F (0H-0T). The other study also includes the data of four groups, three of which were tested under 23 hr. of food deprivation but under different levels of water deprivation; Groups A (23H-0T), B (23H-12T), and C (23H-23T); and Group F (0H-0T). This mode of presentation was adopted because it most clearly indicates the effect on performance of increasing hunger while thirst is held constant in the one case and of increasing thirst while hunger is increased in the other.

The performance medians of the 23-hr. water-deprived groups plus the 0H-0T group (Groups E, D, C, and F) on Day 2 and Trial Block 10 are given in Fig. 1. These two periods were

selected because it was felt that they would be most sensitive to the effects of the treatments. Since acquisition of the hurdle-jump response must be established before the effects of the irrelevant drives can be evidenced in the speed of hurdle jumping, the early trials of Day 1, wherein the hurdle-jump response is less likely to be well established, will not as clearly reflect the effects of the addition of the irrelevant drives. The obtained rank order of performance is exactly that which was predicted, i.e., the level of augmentation of the fear-motivated response is a direct function of the level of deprivation. Since the predictions were made with respect to the relative order of treatment effects, a Jonckheere (1954) "distribution-free k-sample test against ordered alternatives" was performed on the performance measures of Day 2 and Trial Block 10. In both cases, the obtained order of performance was consistent with the predicted order. The data of Day 2 gave an S of 364 which when converted, resulted in a t (134) of 2.577, $p < .025$. The data of Trial Block 10 gave an S of 400 which when converted, resulted in a t (134) of 2.793, $p < .005$.

The performance medians of the 23-hr. food-deprived groups (Groups A, B, and C) plus Group F (0H-0T) on Day 2 and Trial Block 10 are given in Fig. 2. The obtained rank order of performance was again exactly in keeping with the predicted order. A Jonckheere test of the data of Day 2 resulted in an S of 302, t (134) = 2.167, $p < .025$; Trial Block 10 resulted in an S of 298, t (134) = 2.138, $p < .025$, thus indicating that the obtained order of performance was consistent with the predicted order.

In Fig. 3 the means of the reciprocals of latency of hurdle jumping in blocks of 5 trials are plotted for the

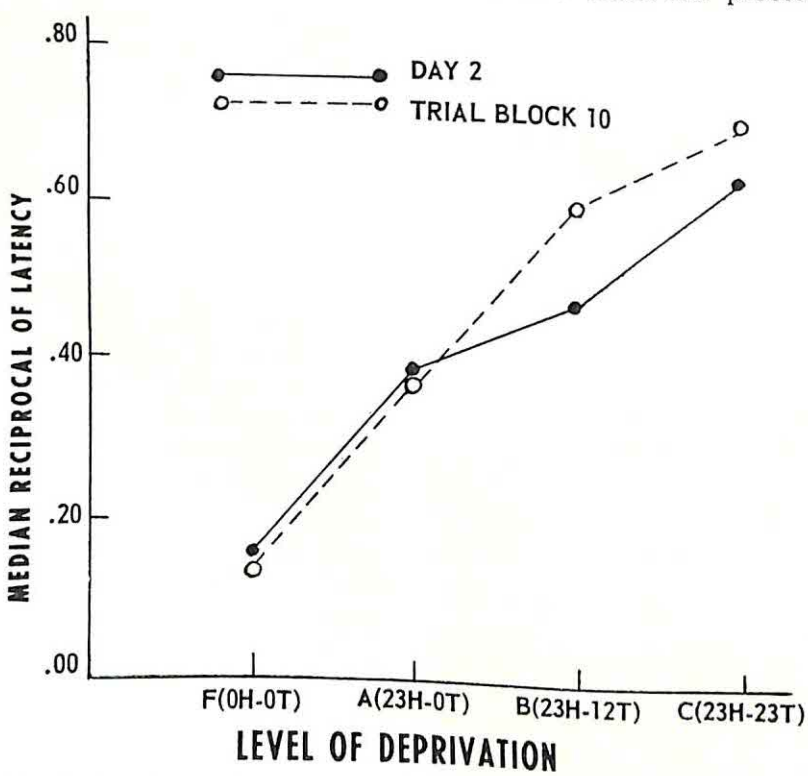


FIG. 2. Hurdle-jumping performance as a function of the addition of thirst to hunger.

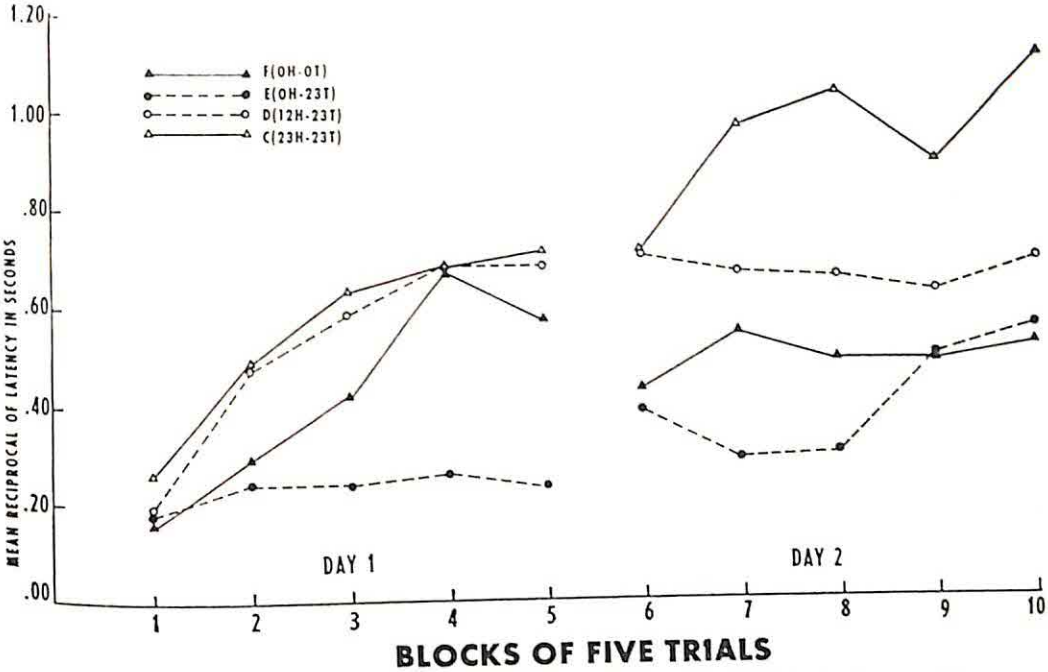


FIG. 3. Hurdle-jumping performance as a function of trial blocks for groups in which hunger has been added to thirst.

groups in which hunger was added to thirst (Groups E, D, C, and F). The summary of a trend analysis of variance over the 10 blocks of trials (Deprivation Conditions × Trials) is given in Table 1. On the basis of the gradients of Fig. 3 it would appear that the significant Trial factor resulted from an overall improvement in performance over the 10 blocks of trials, thus providing evidence that learning occurred. Although the main deprivation effect did not result in a significant mean square, an analysis of the simple Between Groups effect was performed on the means of Day 2 and Trial Block 10, since the interaction mean square was significant. Neither of these analyses resulted in significance.

In Fig. 4 the means of the reciprocals of latency of hurdle jumping in blocks of five trials are plotted for the groups in which thirst is added to hunger (Groups A, B, C, and F). A summary of the trend analysis of variance for these data is given in Table 2. It would appear that the significant Trials factor resulted from an overall improvement in performance over trials, thus indicating that learning occurred.

Discussion

The augmentation of a fear-motivated response through the addition of irrele-

TABLE 1

TREND ANALYSIS OF VARIANCE OF PERFORMANCE OF GROUPS F (OH-OT), E (OH-23T), D (12H-23T), AND C (23H-23T), OVER THE 10 BLOCKS OF FIVE TRIALS

Source	df	MS	F
Between Ss	55		
Deprivation (D)	3	4.8652	2.0604
Error (b)	52	2.3612	
Within Ss	504		
Trials (T)	9	1.2734	13.1686**
T × D	27	.1836	1.8986**
Error (w)	468	.0967	
Total	559		

** $p < .01$.

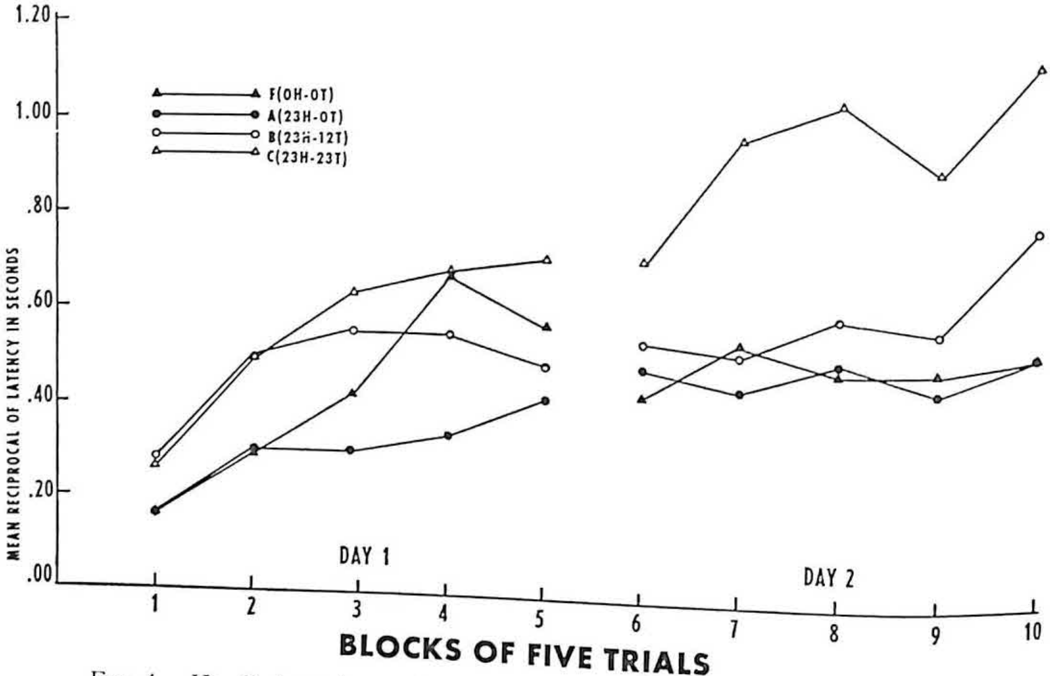


FIG. 4. Hurdle-jumping performance as a function of trial blocks for groups in which thirst has been added to hunger.

vant appetitional needs was substantiated by the data. The results of the Jonckheere tests clearly upheld the predictions generated by Hull's drive-summation hypothesis, that the level of augmentation of the fear-motivated response would be a direct function of the level of deprivation. Within the framework of Hull's theory, it would appear that the addition of various levels and combinations of levels of irrelevant needs summated with the relevant fear drive, thereby increasing D and resulting in heightened performance as evidenced in shorter response latencies.

TABLE 2

TREND ANALYSIS OF VARIANCE OF PERFORMANCE OF GROUPS A (23H-0T), B (23H-12T), C (23H-23T), and F (0H-0T) OVER THE 10 BLOCKS OF FIVE TRIALS

Source	df	MS	F
Between Ss	55		
Deprivation (D)	3	3.3543	1.2547
Error (b)	52	2.6733	
Within Ss	504		
Trials (T)	9	1.1829	11.0551**
T × D	27	.1522	1.4224
Error (w)	468	.1070	
Total	559		

** $p < .01$.

In keeping with the collateral studies of Amsel (1950) and Meryman (1952), the fear-motivated response of the present study was augmented by the addition of irrelevant appetitional needs. Unlike Amsel's study, however, wherein the latency of response of the 22-hr. food-deprived Ss was shorter than that of the satiated group, the present study showed that the performance of the 23-hr. deprived groups (A and E) was not significantly different from the satiated group (F). This difference between Amsel's findings and the findings of the present study is probably the result of a difference in procedure. Amsel shocked his Ss while they were experiencing deprivation, whereas in the present study, Ss were shocked previous to undergoing deprivation. Further, the test day on which Amsel found the significant difference between groups was the day following shock, whereas in the present study, the

crucial test trials did not occur until 8 days after shock. It may be that the heightened performance of Amsel's deprived groups is attributable to an increase in both H and D. Since *Ss* of Amsel's study were deprived during both fear conditioning and testing, the total stimulus complex experienced by his *Ss* would be more similar from training to testing than would that experienced by *Ss* of the present study. Recency would also favor Amsel's procedures, since the considerably shorter delay between training and testing would not provide as much opportunity for the development of competing responses. Considering these differences in procedure it would seem that those followed by Amsel are more likely to show the effect of the addition of one irrelevant need.

With respect to the most severely deprived group, the findings of the present study are exactly opposite those of Kendler (1945). Since Kendler's study involved the use of an instrumental response reinforced solely by food, it may be that his failure to find a motivational effect for his most severely deprived group (22 hr. food deprivation plus 22 hr. water deprivation) resulted from a difference in the effectiveness of the reinforcing agent during acquisition, i.e., the incentive value of the food reinforcement may have been greater for the less severely water-deprived *Ss* than for the 22-hr. water-deprived group, thus offsetting the energizing effects of deprivation. This interpretation is consistent with the data of Bolles (1961) which suggest that when water is restricted, food intake is reduced. Since it is unlikely that hunger and thirst are independent, any study designed to measure the effects of combinations of appetitional needs may be confounded if the dependent variable is either a consummatory response or an instrumental response learned on the basis of thirst or hunger.

Accounting for the results of the present study from Estes' theoretical position presents certain problems. Since the fear response elicited by the CS was learned before *Ss* were subjected to deprivation, i.e. under conditions of satiation, it would appear that the stimulus complex giving rise to fear during test trials should be less similar for the more severely deprived *Ss*. Thus, the more severely deprived *Ss* should be less fearful. If fear is the primary motivational condition for the hurdle-jump escape response, it would seem that Estes' theory would predict that performance should be in an order the opposite from that obtained in the present study. It might be argued, however, that since hurdle jumping is an instrumental response, the additional drive stimuli (S_d) provided by the conditions of deprivation supplement the population of stimuli which can be sampled during the course of acquisition of the hurdle jump, thus allowing for predictions consonant with the data of the present study.

REFERENCES

AMSEL, A. The combination of a primary appetitional need with primary and secondary emotionally derived needs. *J. exp. Psychol.*, 1950, **40**, 1–14.

BOLLES, R. C. The interaction of hunger and thirst in the rat. *J. comp. physiol. Psychol.*, 1961, **54**, 580–584.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

JONCKHEERE, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

KENDLER, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp. Psychol.*, 1945, **35**, 96–109.

MERYMAN, J. J. Magnitude of startle response as a function of hunger and fear. Unpublished master's thesis, State University of Iowa, 1952. Cited by J. S. Brown, *The motivation of behavior.* New York: McGraw-Hill, 1961.

VERPLANCK, W. S., & HAYES, J. R. Eating and drinking as a function of maintenance schedule. *J. comp. physiol. Psychol.*, 1953, **46**, 327–333.

(Received March 14, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 590–595

RECOGNITION MEMORY FOR RANDOM SHAPES AS A FUNCTION OF COMPLEXITY, ASSOCIATION VALUE, AND DELAY[1]

HERBERT JAMES CLARK[2]

University of Illinois

12 groups of 24 Ss each were examined on 2 tests at 3 delays for recognition accuracy of random shapes having 2 levels of complexity and 2 levels of association value (A). A measure of form coding was also taken. Forms of high A were more accurately recognized than forms of low A, and a 2nd recognition test on the same forms resulted in a practice effect for simple but not for complex forms. All delay effects were insignificant. Complex form coding was positively related to A. Simple form coding was infrequent and not related to A. Coding was positively related to recognition accuracy for complex forms only. Conclusions were: (a) Ss probably stored an uncoded image of the entire shape of simple, but not of complex forms; (b) complex forms were presumably remembered by making associations to them.

The purpose of this experiment was to determine the effects of form complexity (Attneave, 1957), form association value (A) (Vanderplas & Garvin, 1959a), and delay between initial and subsequent presentation of a form on recognition accuracy. Four experimental outcomes, based primarily on conclusions arrived at by Deese (1956), were predicted.

1. With increases in delay time, recognition accuracy was expected to decrease more for complex forms than for simple forms, because complex forms contain more information than simple forms.

2. With equally low A's for complex and simple forms, it was expected that, at the shortest delay, complex forms would be more accurately recognized than simple forms. Higher accuracy was predicted for complex forms because they are presumably more discriminable than simple forms.

3. With equally high A's for complex and simple forms, at the shortest delay, the two types of forms were expected to be identified equally well. Deese (1956) suggests that "If the observer is able to code forms with previously stored information, the information conveyed in the forms may make little difference in a test of recognition [p. 3]."

4. Forms of high A were expected to be more accurately recognized than forms of low A. Evidence for the incremental effect of association value on memory has been supplied by Vanderplas and Garvin (1959b), Peterson, Peterson, and Miller (1961), Murdock (1961), Bousfield and Cowan (1962), and Hilgard (1962).

METHOD

Subjects.—One hundred and fifty-four female and 134 male students selected from the University of Illinois elementary psychology *S* pool served in this experiment as a course requirement. In all but one group, the sexes were divided about equally.

Materials.—A Kodak Carousel slide projector with a shutter which controlled exposure duration projected the stimuli onto a

[1] Based on a dissertation submitted in partial fulfillment of the requirements for a PhD degree in psychology at the University of Illinois. The thesis was under the direction of C. W. Eriksen.

[2] Now at the Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

TABLE 1

Test Stimuli

	High	Medium	Low
Simple forms			
4 point	2, 8, 17, 18, 21	19, 20, 22, 25, 27	24, 26, 28, 29, 30
6 point	3, 8, 10, 11, 12	9, 15, 16, 18, 22	13, 19, 20, 21, 30
Complex forms			
16 point	1, 2, 3, 4, 5	6, 8, 9, 10, 12	7, 11, 14, 17, 28
24 point	2, 3, 6, 7, 9	5, 8, 11, 12, 21	10, 15, 16, 22, 28

3 × 4 ft. white screen. Image size was such that each could be contained within a circle having a 1-ft. diameter. The projector illumination selector was set at the 500-w. position. Sixty randomly constructed shapes mounted on 2 × 2 in. slides were the stimuli. These shapes had 4, 6, 16, or 24 points, and were selected from a group of 180 uncurved shapes constructed by Vanderplas and Garvin (1959a). Four- and 6-point shapes are referred to here as simple forms, while 16- and 24-point shapes are called complex forms. The forms selected are listed in Table 1 by their Vanderplas and Garvin identification number.

There were 30 complex and 30 simple forms. Within each category there were 10 forms each of high, medium, and low A. The A measures were taken from Vanderplas and Garvin (1959a), and represent the percentage of 50 *S*s who gave an associative response to a 3-sec. visual presentation of the form.

An attempt was made to match the A's of the 4-point forms with the 6-point forms, and the 16-point forms with the 24-point forms. An effort was also made to match the means and variances of simple and complex forms of high, medium, and low A, respectively. The obtained values are presented in Table 2.

Procedure.—Twelve groups of 24 *S*s each were examined for recognition accuracy in an analysis of variance design which included three levels of delay (5, 10, and 20 min.), two levels of form complexity (simple-complex), and two levels of A (high-low). The *S*s within each group were administered two recognition tests.

Each group was tested separately during approximately a 1-hr. session. The *S*s sat in a classroom 10–20 ft. from the screen. All sessions had two major phases. During Phase I, *S*s viewed 10 forms—each form presented once—read a newspaper during a delay period, and then attempted to identify the 10 learned forms in a forced-choice recognition test in which the 10 learned forms were paired randomly with 10 new forms. A questionnaire on the newspaper content was administered following the recognition test. Phase II was a replication of Phase I with three exceptions: (*a*) The same 10 forms were presented, but in a different random order. (*b*) A different newspaper and questionnaire were used. (*c*) The learned and unlearned forms were paired according to a different random order.

Prior to Phase I, all *S*s participated in a brief practice session which acquainted them with the experimental task. The practice slides were of CVC trigrams. A reward of $2.00 was given to the one *S* within each group who scored highest on the two questionnaires.

The 10 slides to be learned were presented one at a time, each for 0.5 sec., with a 3.5-sec. interval between slides. The *S*s were instructed to look at the forms and remember them. The order of slide presentation was random with the restriction that 4- and 6-point forms were presented alternately as were 16- and 24-point forms. The same random order was used for all groups to hold

TABLE 2

Mean Association Values and Variances for Simple and Complex Forms of High, Medium, and Low Association Value

	Simple[a]		Complex[b]	
	M	S^2	M	S^2
High	45.20	27.36	45.20	24.96
Medium	39.20	10.56	38.80	7.36
Low	35.20	19.36	35.20	20.16

[a] 4 and 6 points.
[b] 16 and 24 points.

TABLE 3

ANALYSIS OF VARIANCE OF RECOGNITION SCORES

Source	*df*	*MS*	*F*
Between *Ss*	287	2.94	
Association value (A)	1	105.06	40.09**
Complexity (C)	1	0.25	
Delay (D)	2	0.22	
A × C	1	3.06	
A × D	2	0.19	
C × D	2	2.82	
A × C × D	2	2.11	
Error (between)	276	2.62	
Within *Ss*	288	1.39	
Phase I (vs.) Phase II (P)	1	11.10	8.47**
P × A	1	2.01	
P × C	1	11.01	8.40**
P × D	2	1.05	
P × A × C	1	0.07	
P × A × D	2	1.65	
P × C × D	2	1.65	
P × A × C × D	2	3.83	
Error (within)	276	1.31	
Total	575	2.16	

** $p < .01$.

constant the interaction between A and serial position.

On each of the 10 recognition test trials, *S* saw two forms, one at a time, each for 0.5 sec., with a 3.5-sec. interval between them. During a 7.5-sec. interval between pairs, *S* designated by writing the number "1" or "2" whether the first or second form presented was the learned form. On half of the 10 trials the first form was a learned form, and on the other half, it was an unlearned form of the same complexity level but of *medium* A. Pairing was done so that the associative relationship between members of a pair was the same for all groups. The order of presentation of the 20 test forms was such that the order of reoccurrence of the learned forms was the same for all groups. The orders of reoccurrence were not the same for the Phase I and Phase II recognition tests.

A questionnaire administered at the end of Phase II inquired of the extent to which *Ss* coded the forms as an aid in remembering them. The question asked was: "Did you name or code the forms, in any way, in order to help you remember them? For example, some people have used names like 'stairway,' 'bird flying,' or 'witch's hat' to identify particular forms." Four alternative answers were provided, and *S* circled one. The alternatives were: (A) always, (B) frequently, (C) seldom, (D) never. The question was taken from Deese (1956, p. 14), and was included to determine the relationships between recognition accuracy and coding. Positive relationships were expected.

RESULTS

Analysis of recognition scores.—The maximum score for the recognition test of each phase was 10. A response was judged correct and scored 1 if the learned form was correctly identified. The total number of correct responses each *S* made was counted and the sums used as scores in a $2 \times 2 \times 3 \times 2$ analysis of variance. The between-group variables were complexity, A, and delay. The within-group variable was Phase I (vs.) Phase II recognition accuracy. Table 3 summarizes the analysis.

Listed in Table 4 are the Phase I and Phase II mean recognition scores for all 12 experimental groups. Also shown for each group are the means and variances of the combined Phase I and Phase II scores. Bartlett's (Edwards, 1960) test for the significance of differences among the variances was not significant. Groups 1–6 viewed complex forms, and Groups 7–12 viewed simple forms. The letters H and L in Table 4 indicate

TABLE 4

MEAN CORRECT RECOGNITION SCORES AND VARIANCES

Group	*M* I	*M* II	*M* I, II	S^2 I, II
Complex				
1. H 5	8.20	8.29	16.50	4.69
2. H 10	8.66	8.41	17.08	3.03
3. H 20	8.08	8.58	16.66	4.14
4. L 5	7.45	7.58	15.04	10.56
5. L 10	7.29	7.50	14.79	5.21
6. L 20	7.54	6.87	14.41	4.42
Simple				
7. H 5	8.29	8.54	16.83	7.44
8. H 10	7.45	8.75	16.20	3.47
9. H 20	8.04	8.54	16.58	5.12
10. L 5	7.33	7.70	15.04	5.86
11. L 10	7.12	7.66	14.79	5.12
12. L 20	7.58	7.95	15.54	4.08

TABLE 5

Relationship between Coding and Association Value for Simple and Complex Forms

	Simple Forms		Complex Forms	
	SH	SL	CH	CL
"Always" and "Frequently"	21	19	48	26
"Seldom" and "Never"	51	53	24	46

whether a group was tested on forms of high or low A. The numbers 5, 10, and 20 denote delay times.

The results of the analysis of variance indicate that: (*a*) All delay effects were insignificant. (*b*) There was no evidence that complex forms were more discriminable than simple forms. At the 5-min. delay, complex and simple forms of low A were identified equally well. (*c*) At the 5-min. delay, complex and simple forms of high A were identified equally well. Since the effects of complexity were insignificant under both the high and low A conditions, A did not wash out the effect of complexity only under the high A condition as was predicted. (*d*) Within both complexity categories, forms of high A were recognized more accurately than forms of low A, thus confirming Prediction 4. A practice effect is indicated by the significant main effect of Phase I (vs.) Phase II recognition accuracy. Investigation of the significant P $\times$ C interaction attributed the effect to simple forms only.

Analysis of coding scores.—The relationship between coding and A is depicted in Table 5. In order to eliminate cells with low frequencies, the coding alternatives "always" and "frequently" were combined. For the same reason, *S*s were combined over delays, resulting in a four-category *S* classification: complex high (CH), complex low (CL), simple high (SH), and simple low (SL).

The chi-square value for a test of independence between simple form coding and A was 0.03 ($df = 1$), indicating no relationship between coding and the Vanderplas and Garvin association values for simple forms. A one-sample (Siegel, 1956) chi-square test showed that simple forms were coded "seldom" and "never" to a significantly greater degree than they were coded "always" and "frequently"; $\chi^2\ (1) = 28.44$, $p < .001$. There was a positive association between complex form coding and A; $\chi^2\ (1) = 12.26$, $p < .001$.

The relationship between coding and recognition accuracy was measured by correlating *S*s' coding scores with their Phase I plus Phase II recognition scores. The four coding alternatives were scored A = 4, B = 3, C = 2, and D = 1. The obtained Pearson product-moment correlation coefficients are presented in Table 6, along with mean recognition scores, mean coding scores, and standard scores which denote the significance of the correlations. These statistics were computed after combining *S*s over delays, and are therefore based on the performance of 72 *S*s. The *z*

TABLE 6

Mean Recognition and Coding Scores, and Correlations between *S*s' Coding and Recognition Scores

Groups	Mean Recognition	Mean Coding	r	z
CH	16.75	2.69	0.49	4.45**
CL	14.75	2.06	0.31	2.81**
SH	16.54	1.91	0.22	2.00
SL	15.12	1.76	0.20	1.81

** $p < .01$.

scores are transformations of the correlation coefficients, and indicate the significance of the correlations when referred to the standard normal distribution (McNemar, 1962). Values over 2.58 were accepted as significant. The relationship between coding and recognition accuracy was significant for complex forms, but not for simple forms.

Discussion

Recognition for *complex* forms was positively related to A and coding, and the latter two variables were positively related to each other. Recognition for *simple* forms was positively related to A, but A was not significantly related to coding, nor was coding significantly related to recognition accuracy. Ignoring the coding data, it appears that *S*s were able to make associations to most forms of high A, and were consequently able to recognize these forms more accurately than forms of low A.

The relationship between coding and A for complex forms suggests that our coding measures and the Vanderplas and Garvin A measures are very similar. The same relationship obtained for simple forms suggests that the two measures are not the same. Other evidence which suggests that our coding data and the Vanderplas and Garvin measures are not the same is the finding that form complexity and coding were directly related. Vanderplas and Garvin (1959a), and Goldstein (1961), found an inverse relationship between complexity and A.

If our coding measure is reliably different from the A measure, it may reflect different processes underlying the perception of simple and complex forms. If such is the case, the two processes are probably not equally efficient. A t test on the difference between the Phase I simple and complex recognition means was insignificant, but a corresponding test on the Phase II means gave a t of 2.370 ($p < .01$). The memory traces established for the simple forms seen during Phase I might have been more stable than those established for the complex forms, resulting in the practice effect. Alternatively, learning may have been equal, but the retroactive effect of the Phase I recognition test may have been greater for complex than for simple forms.

Postexperimental conversations with *S*s suggested one way in which the perception of 4- and 6-point forms may differ from the perception of 16- and 24-point forms. The *S*s were well aware of their making associations to complex forms, and found such coding necessary for accurate recognition. Simple forms were difficult to code, however, and *S*s claimed they relied on remembering an uncoded image of the entire contour of the form. Perhaps the contours of simple forms can be stored, while the contours of complex forms cannot be stored. The practice effect for simple forms, then, may be related to this supposition. It may be easier to build upon a stored image of an entire form rather than on one of several possible associations to a form.

The obtained difference in recognition accuracy between simple forms of high and low A remains unexplained. Perhaps the simple forms of high A had characteristics which allowed images of their contours to be remembered more efficiently than the contours of simple forms of low A.

Finally, the absence of any delay effects suggests that the burden placed on the memory process was well within *S*s' short-term memory, or that the recognition task employed was too gross to detect any memory changes. Had a recall measure of remembering been employed, or had longer or shorter delays been sampled, delay effects might have been found.

References

Attneave, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53**, 221–227.

Bousfield, W. A., & Cowan, T. M. Immediate memory spans for CVC trigrams. Technical Report No. 37, January 1962, Contract Nonr 631 (00), Office of Naval Research.

DEESE, J. Complexity of contour in the recognition of visual form. *USAF WADC tech. Rep.*, 1956, No. 56-60.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

GOLDSTEIN, A. G. Spatial orientation as a factor in eliciting associative responses to random shapes. *Percept. mot. Skills*, 1961, **12**, 15–25.

HILGARD, E. R. *Introduction to psychology.* New York: Harcourt, Brace, 1962.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1962.

MURDOCK, B. B., JR. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625.

PETERSON, L. R., PETERSON, M. J., & MILLER, A. G. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

VANDERPLAS, J. M., & GARVIN, E. A. The association value of random shapes. *J. exp. Psychol.*, 1959, **57**, 147–154. (a)

VANDERPLAS, J. M., & GARVIN, E. A. Complexity, association value, and practice as factors in shape recognition following paired-associate training. *J. exp. Psychol.*, 1959, **57**, 155–163. (b)

(Received March 19, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 596–608

DISCRIMINABILITY AND PREFERENCE FOR ATTRIBUTES IN FREE AND CONSTRAINED CLASSIFICATION[1]

SHIRO IMAI AND W. R. GARNER

Johns Hopkins University

2 experiments on classification were carried out with stimuli consisting of a pair of dots which could vary in Position, Distance between dots, and Orientation of the dots. In free classification, stimulus sets are presented to S, who dichotomously classifies according to any attribute he chooses. In this task choices are affected by individual preferences for attributes as well as by discriminability of the attribute chosen and discriminability of competing attributes. In constrained classification, S is required to sort stimulus sets according to the attribute specified by E, and speed of sorting is measured. In this task, only discriminability of the criterion attribute affects performance.

In a classification experiment, stimuli are generated by combining the different possible levels of two or more attributes to provide a set of multivariate stimuli. The task of *S* in such experiments is to place each stimulus from the total set into one of a number of different classes, either by a differential response or by actually placing stimuli in different groups. Whatever the particular experimental procedure, *S* must demonstrate both differentiation and generalization—differentiation because he must use a different label or group for each class of stimuli, and generalization because he must use the same label or group for all members of the same class.

The purpose of the present experiments is to investigate some of the factors which affect classification performance, in particular those which relate to differences in the attributes which define the stimuli:

Preference for attributes.—While there are many ways of forming classes of stimuli from the total set, the simplest and by far easiest for *S* to use (Shepard, Hovland, & Jenkins, 1961) is for each level of one attribute to define one class. With classes formed from a single attribute, the task of *S* is to use that one (relevant) attribute for differentiation, and to ignore differences in all other (irrelevant) attributes, i.e., he must generalize across different values of the nondifferentiating attributes.

It seems reasonable that some attributes should lend themselves more naturally to such differentiation than other attributes, perhaps because they "stand out" more, are more "salient," or are just preferred by *S* for any of a number of possible reasons. As reasonable as it may seem, however, the determination of these properties of attributes in classification tasks is difficult because of their interaction with other properties such as discriminability. Archer, Bourne, and Brown (1955), for example, suggest that any differences between attributes in ease of concept learning are probably due to differences in discriminability, specifically referring to such differences as found by Heidbreder (1946) and Baum (1954).

[1] This research was done in part under Contract Nonr-248(55) and in part under Contract Nonr-4010(03) between the Office of Naval Research and the Johns Hopkins University. This is Report No. 6 under the latter contract.

(See Hunt, 1962, pp. 55 f., for a fuller discussion of this problem.)

Discriminability of an attribute.—In order for a particular attribute to provide the basis for differentiation of classes, its two or more levels must be discriminable from each other. Furthermore, the different levels on one attribute may be more discriminable than the levels on another attribute. It does not seem reasonable to ascribe to an attribute a special property such as salience, dominance, or even just preference, unless we can be assured that the property is not simply the result of the discriminability of the attribute.

This problem is clear in its extreme. We would not say an attribute is not preferred, or is ineffective, if its two levels cannot be discriminated. In such a case, the attribute is simply not functional. But if two attributes are compared for effectiveness, and one of them has levels which are much more discriminable than the other, we should still be hesitant to say that the more discriminable attribute was the preferred or more effective attribute, because the conclusion could easily be reversed if different levels were used on the variables. Thus we cannot really investigate the problem of the special role of attributes in classification without also investigating the role of discriminability of attributes.

Types of classification task.—Typically, classification tasks have required S to learn a classification rule specified by E, through a series of examples. In a sense, however, classification tasks can be considered to be on a continuum in which the freedom allowed S in his task is varied. At one end of this continuum we would simply give S the entire set of stimuli and ask him to form classes. In such a task, which we shall call *free classification*, the experimental result is the determination of the actual mode of classification used by S.

At the other end of this continuum, the task is completely specified by E and is told to S, so that S has no choice, even initially, about how he is to form his classes. In such a task, which we shall call *constrained classification*, we cannot measure how S forms his classes, but must somehow measure the efficiency with which S can use the different rules specified by E.

It seems quite possible that the role of a particular attribute in classification tasks will itself depend on the nature of the task. Such factors as preference for an attribute might be operative with free classification, where some choice exists for S, but might be of no consequence where S has no choice about his classifications.

Thus these experiments are designed to show the interrelations between preference for attributes, discriminability of attributes, and the amount of freedom allowed S in his classification task.

The Stimuli

Each stimulus consisted of a white card, 10 cm. square, on which was placed a pair of dots, each dot having a diameter of 2 mm. Different sets of these stimuli were generated, and in each set the stimuli could differ from each other with respect to three dichotomous attributes of the pair of dots. In addition, the two levels of each attribute differed in relative discriminability, as determined by the amount of physical difference between the two levels of the attribute.

Stimulus Characteristics

A stimulus characteristic consisted both of a particular attribute of the stimuli and also of its relative discriminability. Each of the three attributes was used with four different relative discrimi-

nabilities, so that there are altogether 12 different stimulus characteristics.

Position of dots (P).—One attribute involved the lateral position of the pair of dots, each pair always being centered with respect to vertical location. When position was not functional as an attribute for a particular set of cards, the pair of dots was also centered horizontally. When it was functional as an attribute, then the pair of dots would be located to the right of center for half the stimuli and to the left of center an equal amount for the other half of the stimuli. For a relative discriminability of P1, the midpoint of the pair of dots was either 2 mm. to the right or to the left of center; for P2, 4 mm.; for P3, 8 mm.; and for relative discriminability of P4, 16 mm. to the right or left of center.

Distance between dots (D).—The second attribute involved the distance between the pair of dots. When distance was not functional as an attribute, the two dots were 10 mm. apart at their centers. For the four relative discriminabilities, the two distances were, respectively: D1, 9.5 or 10.5 mm.; D2, 9.0 or 11.0 mm.; 8.0 or 12.0 mm.; and D4, 6.0 or 14.0 mm.

Orientation of dots (O).—The third attribute was the angular orientation of the dots. When orientation was not functional as an attribute, the pair of dots was always vertical. For the four relative discriminabilities, the dots were rotated clockwise or counterclockwise in the following amounts: O1, 1.5°; O2, 4.7°; O3, 14.5°; and O4, 45°.

Experiment I: Free Classification

Method

Subjects.—A total of 24 male undergraduates was randomly selected from the elementary course in psychology at Johns Hopkins University. Each *S* participated singly in all phases of the experiment, which required two 1-hr. sessions.

Stimulus sets.—All possible sets involving four stimuli (two attributes) and eight stimuli (three attributes) were used. There are three pairs of attributes, PD, PO, and DO, and each attribute of each pair has four relative discriminabilities, which were used in all possible combinations. So there were 16 sets for each pair of attributes, making 48 sets of two-attribute stimuli.

All relative discriminabilities were also used in all possible combinations for the three-attribute stimuli, making $4 \times 4 \times 4 = 64$ stimulus sets of eight stimuli each. Thus there was a total of 112 sets of stimuli used in this experiment.

Task.—The task of *S* was to classify or sort each set of stimuli into two groups of equal size. The *S* was informed of the general nature of the experiment, and the stimuli were demonstrated to him. He was also informed that there was no correct method of classification, so that he was to classify only in a way which seemed reasonable to him.

The recorded measure was the nature of the classification, which could ordinarily be described simply by specifying the attribute which was used as the basis of differentiation.

Stimulus presentation.—Since the experimental measure obtained in this experiment was a free choice of method of classification, we attempted to minimize artifactual influences in the choices, such as could occur either due to the temporal sequence of the sets or to physical arrangement of a particular set.

Both four- and eight-stimulus sets were intermixed in the sequence, such that at the end of each 28 trials, 12 four-stimulus and 16 eight-stimulus sets had been used. In addition, similar sets were kept apart in the sequence, and each *S* had a different order of presentation, arranged to counterbalance for the order with which particular attributes were used.

When a four-stimulus set was presented, it was arranged before *S* in a square so that a vertical separation of the stimuli corresponded to one attribute's dichotomy, and a horizontal separation corresponded to the other attribute's dichotomy. The attributes were assigned to the horizontal and vertical positions in a counterbalanced manner.

The arrangement with eight stimuli was of necessity somewhat more complex. The stimuli were placed before *S* in two rows of four each, and two of the attributes could be selected by selection of columns, but the third attribute had to be selected by selection of two stimuli from each row, not in the same column. Again, the important consideration is that the different attributes were assigned to the different positions in an order counterbalanced both within and between *Ss*.

Results

Preference for attributes.—Since each attribute was paired equally often

TABLE 1

Percent Choice of Attribute by Three Types of *S*s in Free Classification

S Type	*N*	Attribute Chosen			
		Distance	Orientation	Position	Other
D preferring	12	69.0	22.3	5.3	3.4
O preferring	8	22.8	61.9	12.2	3.1
DO preferring	4	39.5	40.0	19.0	1.5
All *S*s	24	48.7	38.4	9.9	3.0

Note.—These percentages represent choices from all two- and three-attribute sets.

with all other attributes, in both two- and three-attribute sets, and with all relative discriminabilities, the overall percentage of times that each attribute served as the basis of the free classification can give some indication of the average preference for each of the three attributes. These average percents, for all *S*s, are shown in Table 1, and indicate that Distance was chosen nearly half the time, Orientation over a third of the time, and Position less than a tenth of the time. Classifications not based on a single attribute occurred only 3.0% of the time.

Experiment II, reported later, shows that the relative discriminabilities of the three attributes are all nearly equivalent in terms of speed of discrimination, a fact which allows us to conclude that these differences in percent choices of attribute do reflect true differences in preference for the attributes.

Individual differences in preference.—Even though there exists an overall difference in preference for the three attributes, there are quite striking individual differences in the pattern of these preferences. Table 1 presents these individual differences in one possible way. The total choices for each *S* were determined, and if an *S* chose one of the three attributes more than 50% of the time, he was labeled as preferring that attribute. Twelve of the *S*s chose Distance more than half the time, and the pattern of all their choices is shown in Table 1. Eight *S*s chose Orientation more than half the time, and their choices are also shown. Four *S*s chose no single attribute more than half the time, but the pattern of their choices makes it clear that they preferred either Distance or Orientation to Position, and they have been so labeled.

It is of particular interest that Position was chosen so seldom. The strong lack of preference for Position as an attribute can be shown also by considering the order of preference of attributes for each individual *S*. With three attributes, there are six possible orders of preference. The two orders with Position most preferred never occurred. The preference order of D > O > P occurred 13 times; the preference order of O > D > P occurred 9 times, and the other two orders occurred once each. In other words, Position was the least preferred attribute by 22 of the 24 *S*s, and it was second least preferred by the other 2.

The individual differences in preference, then, are concerned almost exclusively with whether Distance or Orientation is the most preferred attribute, and all but four *S*s are easily placed in one or the other of these two types. Additional data analyses will be presented only for the D-preferring

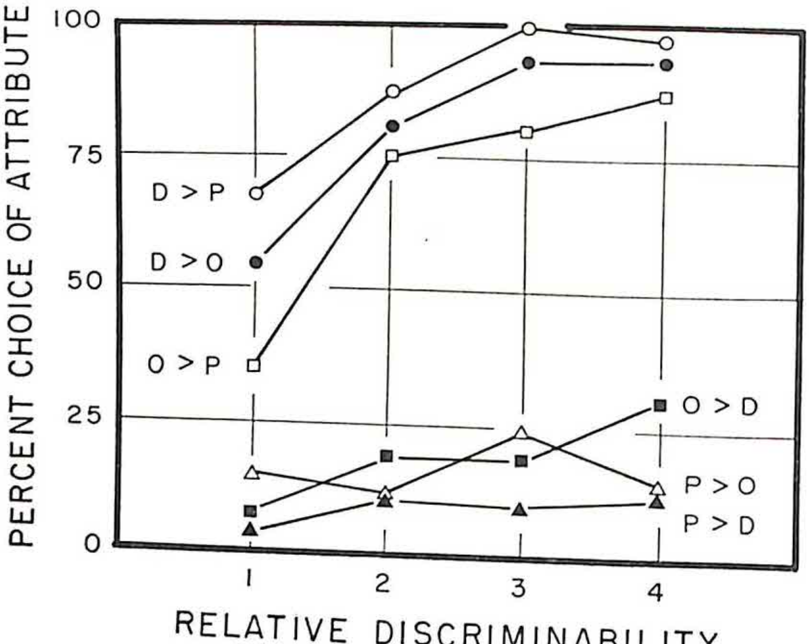


FIG. 1. Effect of relative discriminability on choice of an attribute in free classification. (The abscissa indicates the relative discriminability of each attribute as paired with each other attribute and shown in the figure. The ordinate represents the proportion of choices of the indicated attribute over all discriminabilities of the other attribute. Data from D-preferring *S*s, two-attribute sets only. Note that relative discriminability is defined by the physical difference between the levels of an attribute, as indicated in the text.)

*S*s, but the pattern of results is the same for the O-preferring *S*s, except for the consistent change in preference.

Relative discriminability.—The relative discriminability of an attribute does have a striking effect on its probability of choice as the basis of classification, as shown in Fig. 1. In this figure only choices from pairs of attributes are shown, and there is one curve for each possible choice. For each curve, the relative discriminability of the chosen attribute is shown on the abscissa, and the ordinate shows the percentage of choices averaged over all relative discriminabilities for the attribute not chosen.

These curves group themselves into two groups of three. The top three curves are for the three pairs in which the chosen attribute is also the preferred attribute, as measured by total choices. The bottom three curves are for the pairs in which the nonpreferred attribute is chosen, and these choices occur much less often. Even for these nonpreferred choices, the relative discriminability does affect the probability of choice of the attribute.

Thus the relative discriminability of the chosen attribute affects its choice, but so also does the relative discriminability of the nonchosen attribute. Figure 2 shows the nature of this interrelated effect. For this figure the data have been combined in another way. For each pair of attributes, the number of times that the preferred attribute was chosen was counted, and the resultant percent figures were averaged for all three pairs of attributes, but with the relative discriminability of both attributes kept separate, as plotted in Fig. 2.

This figure shows that an increase in the relative discriminability of a

preferred attribute does increase its probability of choice, and this relation holds for all relative discriminabilities of the nonpreferred attribute. On the other hand, an increase in the relative discriminability of the nonpreferred attribute also increases its probability of choice, but in this case it has an effect only when the relative discriminability of the preferred attribute is low. In other words, if the preferred attribute has a high enough relative discriminability, it will be chosen regardless of the relative discriminability of the nonpreferred attribute. The preference for an attribute seems to give it a priority of choice which will operate whenever the discriminability of the preferred attribute is high enough.

Number of competing attributes.—If a set of stimuli differs in only one attribute, then classification will occur with respect to that attribute. If other attributes are added to the stimuli, they can be considered to be competing attributes in that they will, given sufficient preference and discriminability, interfere with the choice of the original attribute as the basis of classification. Figure 2 can be interpreted as showing the effect of amount of interference produced by one competing attribute with different relative discriminabilities. And that figure shows that if the preferred attribute has a low enough discriminability, and the competing attribute a high enough discriminability, then the nonpreferred attribute can interfere to the extent of being chosen more often than the preferred attribute.

When choices are made from among three attributes rather than two, in effect there is an additional competing attribute. The obtained percentage of choice of each attribute when all three attributes are used together is shown in Fig. 3, as a function of the relative discriminability of the chosen attribute. These data show that Distance is still clearly the most

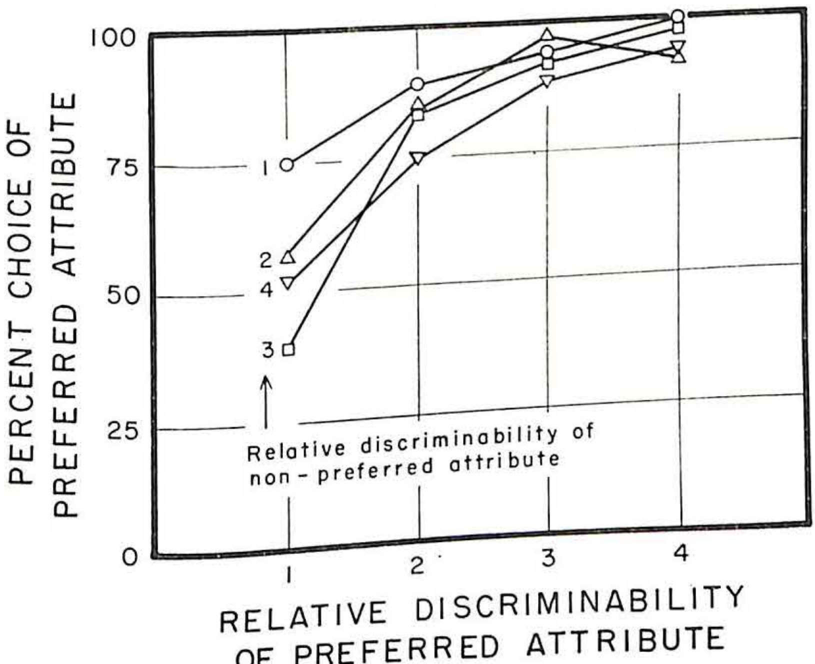


FIG. 2. Effect of relative discriminability of preferred and nonpreferred attributes on choice of attribute in free classification. (Data from D-preferring *S*s, two-attribute sets only.)

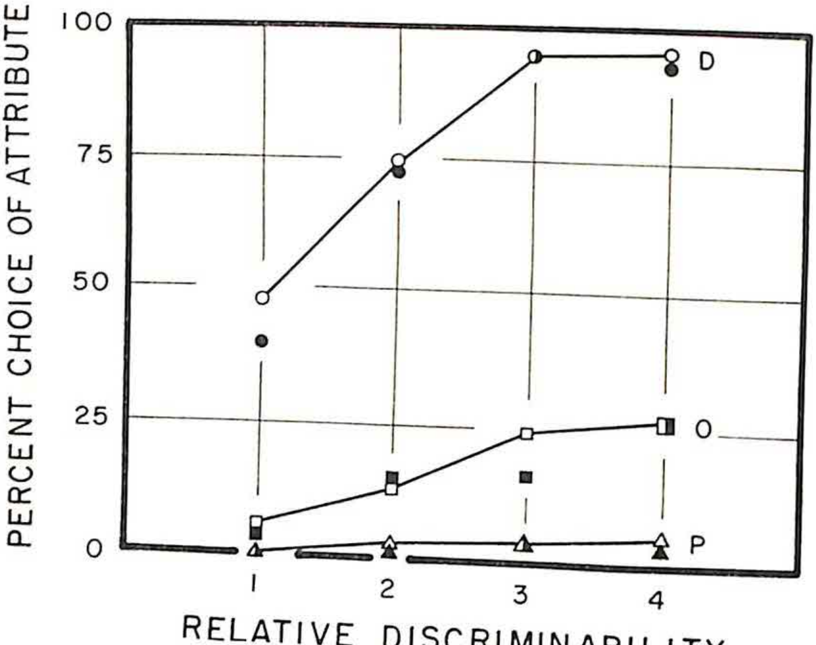


FIG. 3. Effect of relative discriminability on choice of an attribute in free classification with stimuli differing in three attributes. (Data from D-preferring *S*s only. Each curve is for choice of a different attribute as indicated. The open symbols represent actual data, and the closed symbols represent values expected on the assumption of independence of the attributes.)

preferred attribute, with Orientation next most preferred and Position least preferred. Furthermore, an increase in relative discriminability increases probability of choice for all three attributes, as it does for pair choices.

We would like to know whether there is any additional interference effect when all three attributes are combined, over and above the extra interference to be expected by the simple addition of another competing attribute. This question can be answered by comparing the percentage of choices shown in Fig. 3 with values expected on the assumption that the total interference is simply the composite of that produced by the two competing attributes acting independently. These theoretical values are shown as the filled circles in Fig. 3.

The theoretical values were obtained by multiplying the two appropriate probabilities of pair-choices, obtained from Fig. 1, although separately for each *S*. To illustrate, with the composite percents, Distance with relative discriminability of 2 is chosen over Position 87% of the time, and over Orientation 81% of the time. The product of these two values is 70%, which is the value to be expected for choices of Distance with this relative discriminability over all possible combinations of discriminabilities of Orientation and Position.

The rationale for this calculation is as follows: Suppose a single *S* has chosen D2 over P1, P2, and P3, but not over P4; and suppose he has chosen D2 over O1 and O2, but not over O3 and O4. His percentage of choices over the four possible values of P is 75, and his percentage of choices over all possible values of O is 50. Now when D2 is used in a three-attribute set, it is paired with the 16 possible combinations of values

of discriminability of O and P. D will be chosen only if P has a relative discriminability of 1, 2, or 3 *and* O has a relative discriminability of 1 or 2. Only 6 of the 16 combined conditions satisfy this requirement—and this equivalent proportion is obtained by multiplying the two percent figures obtained from the pair choices. The values plotted in Fig. 3 were obtained by carrying this calculation out for each *S* separately, and then averaging. If the pooled data of Fig. 1 are used, the expected values are very slightly smaller.

Overall, the comparison between these theoretical values and the obtained values is very close, and we can conclude that there is no substantial additional interference produced by combining competing attributes, over and above the simple combination of their independent interfering effects. Thus in effect we obtained no additional information from the experiments involving three attributes over that obtained from the pair choices.

Time effects.—At the beginning of the experiment, *S*s are relatively unfamiliar with the attributes, and with their relative discriminabilities. One possible consequence of this lack of familiarity is that preference for attributes would be less important than discriminability.

Over the course of the experiment there was an increase in the relative preference of the preferred attribute. In the first 10 trials, these *S*s chose Distance 60%, Orientation 31%, and Position 9% of the time. In the last 10 trials these figures had changed, respectively, to 79, 19, and 2% of the time. Thus the preferred attribute became even more preferred, and the least preferred attribute was hardly chosen at all at the end of the experiment.

Experiment II: Constrained Classification

Method

Subjects.—Twelve *S*s participated in this experiment, all of them selected from the first experiment. Six of them were Distance-preferring *S*s, and six were Orientation-preferring *S*s, although one of these latter *S*s did not complete the experiment.

Task.—In this experiment the task of *S* was to sort a deck of 32 stimulus cards into two piles differentiated by an attribute specified by *E*. The deck of cards was handed to *S* and he dealt one card at a time off the top and threw it to the right or left. The *S* was told to make no errors, but to work as fast as possible. Both time and errors were recorded by *E*, and the time was told to *S* after each trial.

Stimulus sets.—The stimuli in each deck differed from each other in either one, two, or three attributes. In the latter two cases, sorting was required for each possible attribute, so that each attribute was at times relevant for sorting.

If all attributes are combined in all relative discriminabilities, and sorting is required for each attribute, there are 12 conditions of single attributes, 96 of paired attributes, and 192 of triple attributes, making a total of 300 possible experimental conditions. These were reduced to 150 conditions by never using a relative discriminability of 3 for an irrelevant attribute unless the relevant attribute had the same discriminability, and by never using conditions where the relative discriminability of the irrelevant attribute was considerably less than that of the relevant attribute.

Order of presentation.—A different random series of these 150 conditions was assigned to each *S*. Each series was divided into five groups of 30 decks of stimuli, and each group of 30 was run in 1 hr. An initial practice hour was used, however, in which the decks from the fifth experimental hour were used as practice decks.

Control trials.—An additional deck of cards was used to provide us with a measure of changes in sorting speed either over the course of the experiment or within each 1-hr. session. This control deck consisted of 32 cards on each of which was a 10-mm. circle in the middle. The circle was red for half the cards and blue for the other half, and the two colors were highly discriminable. The *S*s were required to sort these cards by color.

At the beginning of each session, three

TABLE 2

CORRECTED SORTING TIMES FOR THREE ATTRIBUTES AND TWO TYPES OF *S*s

S Type	Attribute			
	Distance	Orientation	Position	*M*
D preferring	6.44	8.25	8.11	7.60
O preferring	8.37	11.85	11.00	10.41
M	7.41	10.05	9.56	9.01

Note.—Each score is the mean corrected sorting time for all sets and *S*s under that condition.

runs on the control deck were used for warm-up, and no measurement was made of time. Then two more measured runs of the control deck were made, followed by 10 experimental decks; then two measured runs of the control deck were made after each 10 experimental decks.

Corrected sorting time.—For each *S*, his control trials were plotted over the course of each session and a smoothed curve drawn. Then the sorting times required for each experimental condition were corrected by subtracting the control time from the obtained time.

These precautions to provide control trials turned out to be somewhat unnecessary, since an analysis of variance of the times obtained with the control trials showed no significant effects either between sessions or within sessions. Apparently the 1-hr. practice session and the warm-up trials stabilized performance. There were, however, large and significant individual differences in the control times, so that the use of the corrected sorting times does give us a reduced inter-*S* variability, and somewhat greater precision for some of the statistical tests.

Results

Effect of preference.—From the first experiment we had established a preference for attribute in free classification, and in this experiment used both Distance-preferring and Orientation-preferring *S*s. Table 2 shows the mean corrected sorting times for these two groups of *S*s, for each attribute. An analysis of variance was carried out on these data, using each *S*'s average score for each attribute as the unit of analysis. This analysis shows that the difference in the mean scores between the two groups of *S*s is not significant, due to the large inter-*S* variances. But more importantly, the interaction between groups of *S*s and

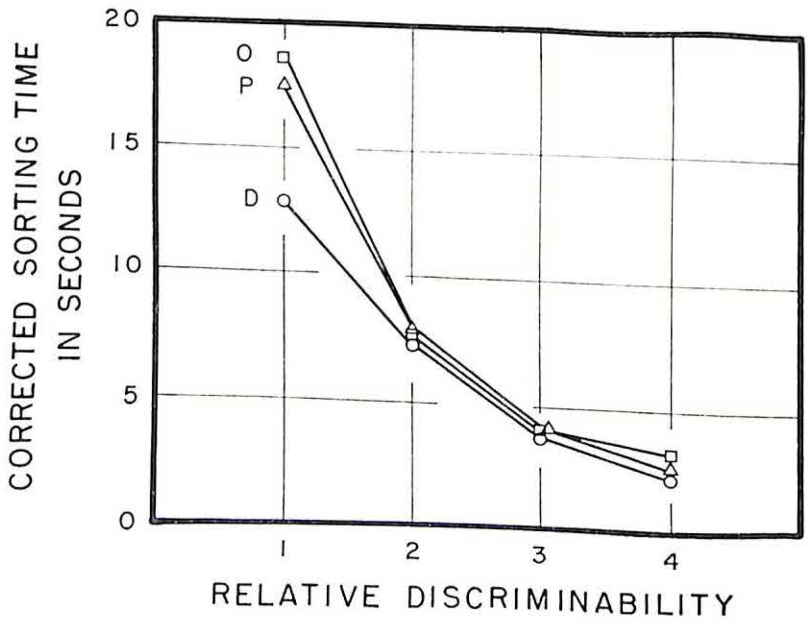


FIG. 4 Effect of relative discriminability on speed of constrained classification for three different attributes, as indicated in the graph. (Each plotted point is the average for all *S*s and all interference conditions.)

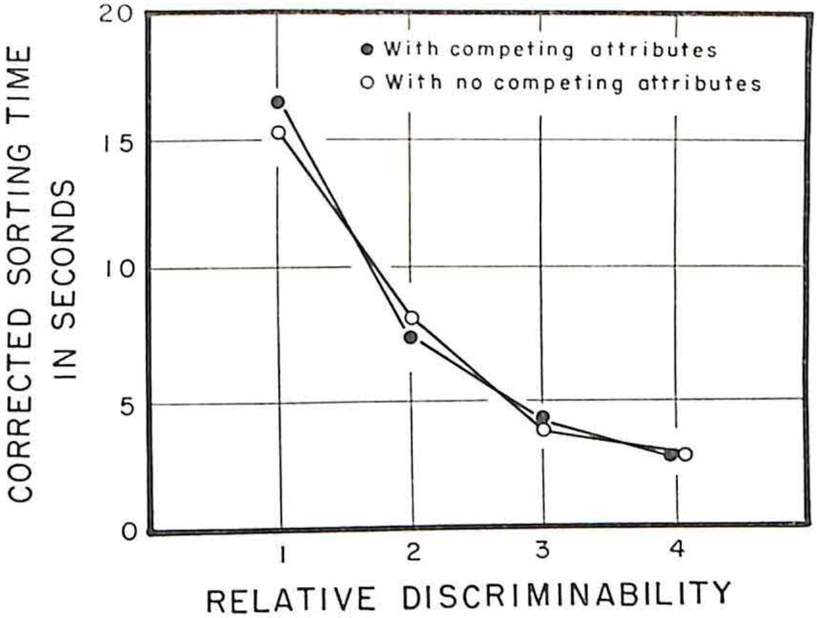


FIG. 5. Effect of relative discriminability on speed of constrained classification with and without the presence of competing attributes. (Each plotted point is the average for all *S*s, attributes, and for both one and two competing attributes.)

attribute does not approach significance. Thus the fact that some *S*s prefer to classify by Distance does not mean that they sort relatively faster by that attribute than *S*s who prefer to classify by Orientation. There is, then, little relation between preference and performance.[2]

On the average, there is a significant effect of attribute, such that Distance provides somewhat faster sorting than do the other two attributes. This result will be shown to be the effect of unequal relative discriminabilities of the attributes.

Relative discriminability.—Figure 4 shows mean corrected sorting time for each attribute as a function of the relative discriminability of the attribute. Since *S* preferences do not affect sorting speed, data for both groups of *S*s and for all conditions have been averaged for these curves.

It is apparent that relative discriminability has a very considerable effect on sorting speed. These curves also show why Distance provides faster sorting on the average than the other two attributes. We had attempted to make each level of relative discriminability to be comparable for all three attributes, and clearly succeeded except for the lowest relative discriminability. Since the differences at the three highest relative discriminabilities are not significant, we cannot conclude that Distance is an inherently better attribute for constrained classification purposes.

Number of competing attributes.—Figure 5 shows mean corrected sorting time as a function of relative discriminability for the cases where there was only one attribute and for all combined cases where there were

[2] It is, of course, logically possible that *S*s' preferences, as exhibited in the first experiment, had disappeared in the time required to initiate the second experiment, but this seems highly unlikely since the preferences were stable over two 1-hr. sessions, and had actually increased in severity over the course of the first experiment.

competing attributes. Analysis of variance shows no significant difference between these two conditions, so that we can conclude that the presence of interfering attributes has certainly little effect at best on constrained classification—a result again unlike that found with free classification.

Error analysis.—The total number of errors made was too small to allow detailed analysis, but what effects appear at all simply confirm the results obtained with time measures. For example, more errors are made with the lowest relative discriminability than with the higher discriminabilities. There is, then, no evidence that conditions giving faster sorting speeds did so by the production of more errors. Thus there was no compensating effect.

Discussion

The major purpose of these experiments was to compare factors affecting performance of two different types of classification task, with the particular intent of determining whether it is possible to determine a special property of some attributes compared to others which will make them preferred. A comparison of tasks for each of several factors will facilitate an understanding of the difference between tasks.

Relative discriminability.—These experiments show that as the difference between the two levels of an attribute increases, that attribute is more apt to be chosen as a basis of classification in free classification, and that sorting on the basis of that attribute is faster. This result for the constrained classification task is not new, since Reed (1951) had shown this result for sorting by attributes of visual area and hue.

The fact that free classification is also severely affected by the relative discriminability of the attributes makes very clear that it is not possible to talk about the properties of a particular attribute without consideration of relative discriminability. Thus it is not sensible to say, for example, that concepts (classifications) can be formed more easily with color than with size unless we know that these two attributes are, in the given situation, equally discriminable.

Competing attributes.—The lack of any effect of number or discriminability of competing attributes on sorting speed is a little surprising, but quite in keeping with previous research. Various *E*s on sorting tasks have found either no difference due to the presence of competing attributes (Archer, 1954; Morin, Forrin, & Archer, 1961), or else very slight differences (Crossman, 1953; Gregg, 1954; Reed, 1951). The lack of effect in our experiment was surprising only in that we used more extreme conditions of interference. For example, some of the sorting conditions required sorting by an attribute with relative discriminability of 1, against two attributes each with a relative discriminability of 4. Even under this circumstance, sorting was not impaired by the presence of the competing attributes.

In the free classification task, the problem of competing attributes is a little different. If a set of stimuli differs on just a single attribute, then classification can occur only with respect to that attribute. But interference must occur if two such attributes are combined into a single stimulus set, since classification cannot occur by both of them simultaneously. However, the interfering effects found with free classification concern much more than this inevitable result, since the nature of the choices is modified by the relative discriminability of the competing attribute. Thus in free classification, the discriminabilities of all attributes enter into the result, whereas in constrained classification only the discriminability of the relevant attribute affects performance.

Preference for attributes.—The logical problem of this paper has been to determine whether we can say that one attribute is preferred over another (or is more salient, easier to use, etc.). We need to know that the attributes being compared are the same with respect to

other properties which could affect the choice, and the property we are most concerned with is that of discriminability of an attribute.

Since in the free classification experiment there were substantial individual differences in choice of attribute, we have some confidence that the preference is not an artifact. It seems unlikely that differences in discriminability would display the same individual differences as the differences in preference, so that the differences in preference are probably real and exist over and above any differences in discriminability.

A major purpose of the constrained classification experiment was to provide a means of determining whether the three attributes were equally discriminable at each of the four levels of relative discriminability we chose. Our results showed that speed of sorting was the same for all three attributes except at the lowest relative discriminability. In order to determine whether this difference at this lowest level of discriminability could have affected the preference measures, we tabulated choices for all cases where none of the attributes had its lowest level of discriminability. This tabulation showed that the preference pattern for individual *S*s remained the same, but the differences in preference within each group of *S*s was stronger. This result is consonant with the relation shown in Fig. 2, which indicates that if the preferred attribute is discriminable enough, it will be chosen regardless of the discriminability of the competing attributes.

Thus if we can accept the speed of sorting as a legitimate measure of discriminability, then we can say that there exist preferences for some attributes over others, and that these preferences are not due to differences in discriminability.

Comparison of characteristics.—Table 3 shows a very useful way of summarizing the differences in these two tasks, and the relative role of preference for attribute and discriminability in these two tasks. Each of the 12 stimulus characteristics (defined by three attributes and four levels of relative discriminability) are ranked in this table according to their effect on performance in each task and for each group of *S*s. For the constrained classification, the ranks are obtained simply by determination of mean sorting time for all conditions in which that stimulus characteristic provided the classification criterion. The correlation between rankings (rho) for the two groups of *S*s in constrained classification is .96, indicating that the type of *S* does not affect sorting performance. But simple inspection makes the predominant role of relative discriminability even clearer, since in no case does an attribute provide faster sorting than another when it has a lower relative discriminability. Ranking with respect to relative discriminability is perfect.

The rankings for free classification were a little more complicated to obtain, since there exist no direct comparisons between levels of relative discriminability for the same attribute. We constructed a pair-comparison matrix for all paired attributes and conditions, and completed the matrix by assuming that if two levels of relative discriminability on the same attribute had been paired,

TABLE 3

RANKINGS OF STIMULUS CHARACTERISTICS IN FREE AND CONSTRAINED CLASSIFICATION FOR TWO TYPES OF *S*s

Rank	Free Classification		Constrained Classification	
	D Preferring	O Preferring	D Preferring	O Preferring
1	D4	O4	D4	D4
2	D3	O3	P4	P4
3	O4	O2	O4	O4
4	D2	D4	D3[a]	D3
5	O3	P4	P3[a]	O3
6	O2	D3	O3	P3
7	D1	P3	D2	O2
8	P4	O1	P2	D2
9	P3	D2	O2	P2
10	O1	P2	D1	D1
11	P2	D1	O1	P1
12	P1	P1	P1	O1

Note.—The letter indicates the attribute and the number following it indicates the relative discriminability. A lower rank number indicates higher frequency of choice or faster sorting.
[a] Tied ranks.

the more discriminable attribute would always have been chosen.

The correlation between ranks for the two groups of *S*s in free classification is .63. If, however, we exchange O and D characteristics for the O-preferring *S*s, to reflect the difference in preference, then the correlation becomes .90. Inspection again shows the great importance of the particular attribute in free classification. To illustrate, Distance-preferring *S*s choose the least discriminable level of Distance over the most discriminable level of Position.

To summarize, in constrained classification only the discriminability of the criterion attribute is important in speed of sorting. In free classification, discriminability of the chosen attribute, discriminability of interfering attributes, and a preference for attribute over and above these factors all affect performance.

REFERENCES

Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Archer, E. J., Bourne, L. R., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

Baum, M. H. Simple concept learning as a function of intralist generalization. *J. exp. Psychol.*, 1954, **47**, 89–94.

Crossman, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice-response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

Gregg, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

Heidbreder, E. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35**, 191–223.

Hunt, E. B. *Concept learning.* New York: Wiley, 1962.

Morin, R. E., Forrin, B., & Archer, W. Information processing behavior: The role of irrelevant stimulus information. *J. exp. Psychol.*, 1961, **61**, 89–96.

Reed, J. B. The speed and accuracy of discriminating differences in hue, brilliance, area and shape. Technical Report No. SDC-131-1-2, September 1951, Mount Holyoke College.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. Learning and memorization of classifications. *Psychol. Monogr.*, 1961, **75**(13, Whole No. 517).

(Received March 23, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 609–612

REACTION TIME TO CHANGES IN THE INTENSITY OF WHITE NOISE [1]

DAVID H. RAAB AND MITCHELL GROSSBERG

Brooklyn College of the City University of New York

Reaction time to a noise burst (ΔI) added to ongoing noise (I) was found to decrease with increasing ΔI and with decreasing I. For constant values of ΔI/I, RT generally decreased with increasing I—a finding of significance for decision models of RT.

The current literature on psychoacoustics includes a number of papers in which signal-to-noise ratio, detectability, loudness, and reaction time are related to each other. The d' index of detectability—as developed by Tanner, Swets, and Green—is equal to $\sqrt{2E/N_0}$ for a pure tone signal known exactly to an "ideal observer" (see Green, 1960b). (E = signal energy; N_0 = noise power density.) When the signal is itself a burst of noise, d' is a monotonically increasing function of the ratio of signal power to noise power (Green, 1960a). Creelman (1963) has shown that tone bursts of equal loudness have identical d' indexes. McGill (1963) has outlined a detection model of simple reaction time (RT) and has inferred from the model that RT and loudness are inversely related and that RT distributions are identical for signals that are equally detectable.

McGill cites Greenbaum's (1963) data in support of the second inference. Greenbaum measured RTs to the onset of a 1,000-cps tone masked by continuous noise. Signal-to-noise ratio (signal level: noise spectrum level) was fixed at approximately 30 db. Similar RT distributions were obtained for four different levels of masking noise.

Actually, as McGill noted, there were differences among the four RT distributions. Median RTs, for example, decreased systematically with increasing mask level; a 10-msec. decrease in RT resulted from a 30-db. increase in masking noise level (see McGill, 1963, Fig. 6).

Equal signal-to-noise ratios (S/N) did not generate equal RTs in Chocholle's (1943, 1944) studies with tonal signals and tonal masks. In all cases, RT was shortened as background stimulation was increased in intensity.

Visual RTs have been measured to increases (ΔI) in the luminance of a portion of a lighted background (I). Decreased RTs at higher levels of I (for constant $\Delta I/I$) can be seen in studies by Piéron (1936), Durup and Piéron (1937), Steinman (1944), and Bartlett and MacLeod (1954).

In the present study, simple RTs were measured to noise bursts added to background noise. Background intensities—all within the region described by Weber's law—covered a range of 60 db.; S/N was varied between 0.63 and 10^7, i.e., between -2 db. and 70 db. Miller (1947) has analyzed the masking of noise by noise and has noted that this instance of masking is operationally identical to intensity discrimination; masking noise (N) and background intensity

[1] This research was supported by a grant (GB-1118) from the National Science Foundation and by funds provided by Brooklyn College. The data were collected by Robert Baron and David Charney as part of an honors course.

(I) are the same, and signal intensity (S) equals increment intensity (ΔI). The effect of wave-form correlation on these relations has been discussed by Pfafflin and Mathews (1962) and by Raab, Osman, and Rich (1963).

Method

Apparatus.—Masking and increment noises were generated and band-limited (100–6,000 cps) in separate channels. They were switched independently and attenuated independently before being mixed to produce the stimuli for each trial. The acoustic output of S's earphone (PDR-10 mounted in a MX/41-AR cushion) was measured by a WE 640-AA microphone in an ASA Type 1 coupler. Sound-pressure levels (SPLs) given in this report are all referred to 0.0002 dynes/cm².

A Berkeley 7350 EPUT counter was used to measure RTs.

Procedure.—Four levels of masking noise (25, 45, 65, and 85 db. SPL) were used. Increment signals were varied from 23 to 95 db. SPL. For the 25-db. mask, therefore, relative increments ($\Delta I/I$) ranged from -2 db. to 70 db.; for the 85-db. mask, $\Delta I/I$ ranged from -2 to 10 db. These stimuli were delivered through one earphone.

Reaction-time trials were presented every 12 sec. Each trial began with the presentation of the masking sound. The increment noise was added to the mask (and the RT counter started) 2.0, 2.25, or 2.5 sec. after the onset of the mask. The counter was stopped when S depressed a normally closed telegraph key. Mask and increment noises were terminated together 3.2 sec. after the onset of the former—half a second or more after S had responded.

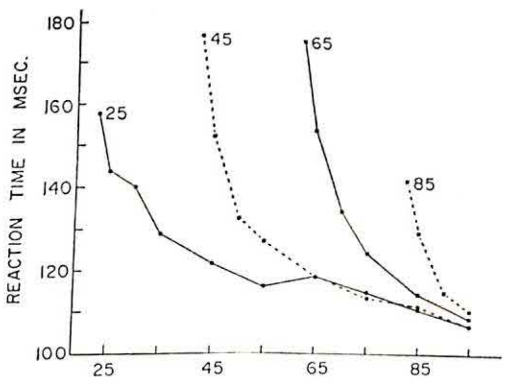


FIG. 1. Reaction time as a function of increment SPL. (The parameter is the SPL of the masking noise. Each data point is the mean for three Ss.)

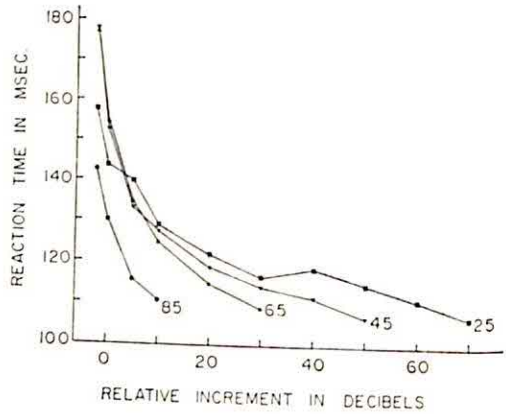


FIG. 2. Reaction time as a function of the relative increment ($\Delta I/I$) in db. (The parameter is the SPL of the masking noise. Each data point is the mean for three Ss.)

Experimental sessions consisted, in general, of four blocks of 28 trials each. Each of the 28 different mask-increment combinations was presented once in a test block. Since E and S changed roles after each block of trials, the duration of each block and the interval between successive blocks (for each S) was approximately 6 min. Four practice trials were presented before the first block of each test session; the three remaining blocks of each session were each preceded by one practice trial.

Subjects.—Three practiced Ss were used. All three also served as Es.

Results and Discussion

Twenty test sessions with each S provided mean RTs based on 80 responses. Standard deviations of the distributions were also computed. They are of the order of 20 msec. and include intertrial variations within a test session as well as day to day variability. Since the principal trends are the same for the three Ss, group means are plotted in Fig. 1 and 2.

Figure 1 shows RT to decrease with increasing increment intensity. The general ordering of the curves shows

that RT is also a function of masking sound pressure. The dependence of RT on both background (I) and increment (ΔI) agrees with the results of the visual and auditory experiments cited earlier.

Of special interest are the relations between RT and $\Delta I/I$. These are given by the plots in Fig. 2. For relative increments of 10 db. or more, the curves are ordered such that RT to a given relative increment is shorter at higher intensities of masking noise. These data confirm—in the case of noise added to noise—what has already been reported (see above) for light added to light, clicks added to noise, and tones added to tone or noise. Equal S/N ratios do not everywhere generate equal RTs.

That RT to a masked stimulus shows a kind of "recruitment" appears, at first glance, to contradict the Detection Theory of Tanner, Swets, and Green. Actually, the finding that *response latency* is not determined solely by S/N is not inconsistent with the view that *detectability* is a function of S/N. The O in a detection experiment is asked to decide about the presence or absence of weak signals in noise. For clearly detectable stimuli ($S \gg N$), on the other hand, loudness judgments, for example, are not specified solely by S/N (Gleiss & Zwicker, 1964; Lochner & Burger, 1961; Raab & Osman, 1962). Creelman's (1963) finding that short tones of approximately equal energy are both equally detectable and equally loud applies only to weak signals ($d' \leqslant 4$).

A similar situation obtains in the case of auditory and visual RT. The present results—together with those from several prior studies—show that RT to fixed ratios of $\Delta I/I$ is shorter at higher intensities of I. If decision models are to be used to describe simple RT to partially masked stimuli, they will have to be different from the detection models currently employed.

REFERENCES

Bartlett, N. R., & MacLeod, S. Effect of flash and field luminance upon human reaction time. *J. Opt. Soc. Amer.*, 1954, **44**, 306–311.

Chocholle, R. Relation de la latence d'une sensation auditive différentielle avec l'amplitude d'une variation brusque d'intensité. *CR Soc. Biol., Paris*, 1943, **137**, 751–752.

Chocholle, R. Étude de la psychophysiologie de l'audition par la méthode des temps de réaction. *Annee psychol.*, 1944, **45**, 90–131.

Creelman, C. D. Detection, discrimination, and the loudness of short tones. *J. Acoust. Soc. Amer.*, 1963, **35**, 1201–1205.

Durup, G., & Piéron, H. Sur la marge réductible des temps de réaction à des accroissements brefs de brillance. *Annee psychol.*, 1937, **38**, 148–161.

Gleiss, N., & Zwicker, E. Loudness function in the presence of masking noise. *J. Acoust. Soc. Amer.*, 1964, **36**, 393–394.

Green, D. M. Auditory detection of a noise signal. *J. Acoust. Soc. Amer.*, 1960, **32**, 121–131. (a)

Green, D. M. Psychoacoustics and detection theory. *J. Acoust. Soc. Amer.*, 1960, **32**, 1189–1203. (b)

Greenbaum, H. B. *Simple reaction time: A case study in signal detection.* (Doctoral dissertation, Columbia University) Ann Arbor, Mich.: University Microfilms, 1963, No. 63–7421.

Lochner, J. P. A., & Burger, J. F. Form of the loudness function in the presence of masking noise. *J. Acoust. Soc. Amer.*, 1961, **33**, 1705–1707.

McGill, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Ed.), *Handbook of mathematical psychology*. New York: Wiley, 1963. Pp. 309–360.

Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. Acoust. Soc. Amer.*, 1947, **19**, 609–619.

Pfafflin, S. M., & Mathews, M. V. Energy-detection model for monaural auditory detection. *J. Acoust. Soc. Amer.*, 1962, **34**, 1842–1853.

Piéron, H. Latence de perception des accroissements de luminosité. *Annee psychol.*, 1936, **37**, 1–16.

Raab, D. H., & Osman, E. Effect of masking noise on lateralization and loudness of clicks. *J. Acoust. Soc. Amer.*, 1962, **34**, 1620–1624.

Raab, D. H., Osman, E., & Rich, E. Effects of waveform correlation and signal duration on detection of noise bursts in continuous noise. *J. Acoust. Soc. Amer.*, 1963, **35**, 1942–1946.

Steinman, A. R. Reaction time to change compared with other psychophysical methods. *Arch. Psychol., N. Y.*, 1944, **41**, No. 292.

(Received March 23, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 613–617

HABITUATION OF ALTERNATION BEHAVIOR[1]

P. E. FREEDMAN[2]

University of Iowa

Alternation behavior of rats in a 2-alternative maze was investigated in 2 experiments. Neither number of immediate prior exposures nor the presence of choice were found to be significant variables. However, brightness was shown to be an important factor in establishing a "chance" performance criterion. The demonstration that alternation habituates over trials was considered in respect to its effect on learning theory.

Alternation behavior, the tendency for an organism, after responding to one alternative of a multiple alternative situation, to respond to another alternative when given a second opportunity, is a well-established phenomenon in animal maze studies. However, the implications of the alternation phenomenon for learning theory generally have been overlooked. Even when the initial exposure is reinforced the tendency to alternate on subsequent trials is significant early in training (Estes & Schoeffler, 1955; Montgomery, 1952; Rothkopf & Zeaman, 1952). A major theoretical question is: as learning takes place, does alternation habituate over trials or does it remain as an opposing tendency?

Two previous studies bear directly on this question. Walker (1956) gave *S*s one pretest exposure (PTE) trial followed by one test trial each day and obtained a level curve of response alternation over a series of 7 days. The level of alternation was significantly greater than 50%. Fowler, Blond, and Dember (1959), using two and eight forced PTE trials per day prior to a test trial, found initially high percentages of alternation with a decline over days, with a somewhat more rapid decline for the eight PTE trial *S*s. At the end of 10 days both curves appear to level off slightly below the 50% alternation level. For the results reported, all trials were reinforced.

Walker's study would indicate that alternation behavior does not habituate. Fowler et al. are equivocal on this point since the decline in alternation could be attributed to a greater habit strength being progressively built up to the PTE brightness.

If the alternation tendency does not habituate, the following assumptions and predictions would be made in a two-alternative situation: (*a*) If *S* is forced to one of the stimuli, e.g., response to white (Rw), on a PTE trial and is reinforced, there is an increment in the habit strength of Rw, and a satiation or exposure factor to the white stimulus (Sw) is developed. Since the satiation factor is initially stronger than the learning factor, there is a greater tendency for *S* to respond to the nonexperienced black stimulus (Sb) on the test trial. If *S* is reinforced for this latter response, an increment in its habit strength results, and both responses would then have equal habit strengths.

[1] This experiment represents a portion of a dissertation submitted to the Psychology Department of the State University of Iowa in partial fulfillment of the requirements for the PhD degree. The author wishes to acknowledge his indebtedness to Kenneth W. Spence for advice and assistance throughout the course of the investigation.

[2] Now at United States Army Natick Laboratories, Natick, Massachusetts.

Repeating this procedure over a period of days, one PTE and one test trial per day, the curve representing percent alternation over days should remain high and stable. (*b*) If two or more forced PTE trials to Sw precede the test trial each day, the difference over days between the habit strengths of Rw and Rb would increase due to different number of increments in the learning factors. Consequently, the habit strength of Rw progressively would become stronger than that of Rb over a series of days, and the percent alternation curve should decline. However, since Rb would still be made occasionally, over extended training the habit strengths of Rw and Rb should approach the same asymptote and hence approach equality. Under this condition the satiation factor would again tend to make *S* respond to Sb, and the curve of percent alternation should rise toward its original level.

The present study tested these predictions in a simple two-alternative maze. An additional variable investigated was the technique of forcing. Studies on alternation varying number of forced trials have employed physical barriers to control the stimulus to which *S* is exposed on a forced trial. The present investigation compared the effects of this forcing technique with a method of controlling stimulus exposure which did not employ a physical barrier.

Experiment I

Method

Apparatus.—A wooden maze was constructed with a $9 \times 2\frac{1}{2}$ in. start box opening into a choice point allowing access to two parallel $42 \times 3\frac{1}{2}$ in. alleys which shared a common wall. The last 12 in. of the alleys were used as goal boxes. The Plexiglas ceiling of the start box was 4 in. high; the ceiling of the remainder of the maze was $9\frac{1}{2}$ in. high.

The start box and choice point were painted neutral gray. The alleys were either black or white, and sets of alleys could be interchanged to present black-white, two black, or two white alleys at the choice point. A metal insert, painted the same brightness as the blocked alley, could be positioned at the entrance of an alley for a forced trial.

Subjects.—One hundred and forty naive hooded rats from the colony maintained by the Department of Psychology of the State University of Iowa served as *Ss*. At the start of the experiment all *Ss* were between 100 and 130 days old.

Maintenance schedule.—The *Ss* were placed on a 22-hr. water-deprivation schedule 10 days prior to the start of the experiment and remained on this schedule throughout the experiment. The *Ss* were given access to water for a 15-min. period beginning 15 min. after each daily experimental session. Food was available in the living cages at all times.

Training procedure.—All *Ss* received one or two PTE trials each day under either a Free Choice (C) or Forced Yoked (Y) condition. Since there was greater initial interest in the C condition, 80 *Ss* were assigned to the C condition, and only 60 *Ss* were assigned to the Y condition. Under the C condition *S* was allowed to choose the right or left alley on a PTE trial, but both were the same brightness, i.e., white-white or black-black. Each *S* in the Y condition was yoked to one of the first 60 of the *Ss* in the C condition and was forced by a physical block to make the same PTE response. Within each PTE condition half the *Ss* had one PTE trial, and the other half had two PTE trials per block. The PTE brightness was black for one half of the *Ss*. After one or two PTE trials each day, all *Ss* were given a test trial where *S* could choose between a black-white alternative. For each *S* black was on the left for half the test trials. All trials were reinforced with .5 cc of water. The *Ss* were retained in the goal box for 15 sec. The daily series of trials for each *S* was run consecutively. The experiment was continued for 40 blocks of trials.

Results

The response of each *S* on each test trial was scored as 1 if *S* alternated and 0 if *S* did not alternate: i.e., chose the same brightness to which he was exposed on the PTE trial(s). The data were analyzed with respect to the

first 10 test trials individually and the last 30 test trials in 5-trial blocks.

Analysis of variance with Method (M), Number (N), and Brightness (B) of PTE trials as between factors and trials (T) as the within factor showed B, T, and B × T interaction significant ($p < .001$) over the first 10 trials. Over the last 30 trials B effects were significant ($p < .001$) and the M × N × B and the T × B interactions reached the .05 level.

Analysis of the first 10 trials within PTE brightness showed that only the trials effect was significant, both within white and within black PTE brightnesses ($p < .005$ and $p < .025$, respectively). There were no significant effects over the last 30 trials in either the white or black PTE brightnesses. The curves of percent alternation for white PTE and black PTE are presented in Fig. 1 and 2.

The apparatus minimized differential right-left kinesthetic cues. No position preferences were evident in *S*s' responses.

Experiment II

The results of Exp. I which indicate that there was no tendency for percent alternation to increase with extended training lead to the tentative conclusion that the alternation tendency did habituate. However, the results do not preclude the possibility that the procedures used in Exp. I did not allow the habit strengths of the two alternative responses to be equalized or approach equality. Furthermore, even though the overall curve for percent alternation approached and leveled off at about the 50% level, the alternation curves within PTE brightness leveled off at about 80% and 20% for white and black PTE, respectively. Therefore, a second experiment was designed in which

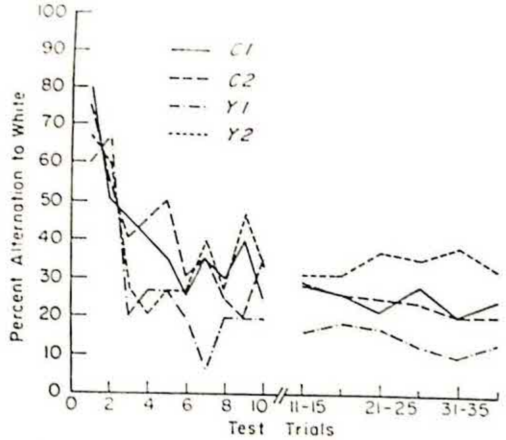


FIG. 1. Percent alternation for the C1, C2, Y1, and Y2 Black PTE groups as a function of trials. (The first 10 test trials are plotted individually. The last 30 trials are plotted in blocks of 5 test trials.)

brightness preference and alternation behavior could be recorded while the learning factor to each brightness was kept approximately equal.

Method

Apparatus.—Same as in Exp. I.

Subjects.—Twenty hooded rats of the same age and origin as those in Exp. I served as *S*s.

Maintenance schedule.—Same as in Exp. I.

Training procedure.—The handling and preexperimental treatment of *S*s was the same

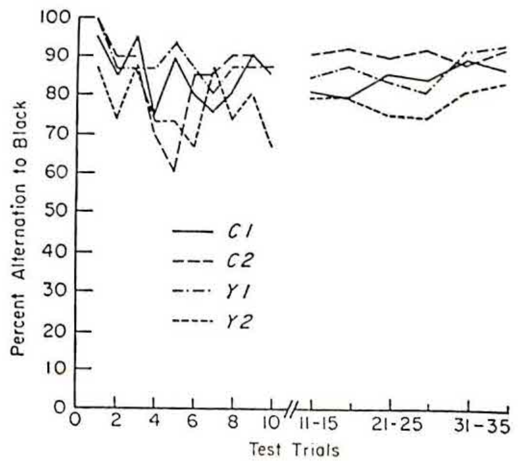


FIG. 2. Percent alternation for the C1, C2, Y1, and Y2 White PTE groups as a function of trials. (The first 10 test trials are plotted individually. The last 30 trials are plotted in blocks of 5 test trials.)

as in Exp. I. All *S*s received the same experimental treatment. Each *S* received four trials per day. On the first two trials *S* was presented with a black-white choice. The left-right positioning of the brightness for these two trials was BW, BW; WB, WB; BW, WB; or WB, BW. Each *S* received a different combination over days. On the last two trials *S* was forced to a particular brightness by a physical barrier. The barrier was painted the opposite brightness from the open alley. The brightness to which *S* was forced was determined by his responses on the first two trials. In each block of four trials *S* experienced each brightness twice. The order and right-left positioning of brightness on the forced trials was determined randomly, but no *S* had all left (or right) turns or experienced each brightness consistently on a particular side in a four-trial block. This training continued for 5 days. All trials were reinforced.

Results

Considering the first trial of each day, of the 100 total opportunities black was chosen 80 times and white was chosen 20 times. Following an initial choice of black, *S*s alternated (chose white) 16 out of 80 times or 20% of the time; whereas following an initial choice of white, *S*s alternated (chose black) 17 out of 20 times or 85% alternation.

Fifty-five percent of the *S*s alternated on the first day. Alternation dropped to 35% on the second day and to 25% on the third, fourth, and fifth days.

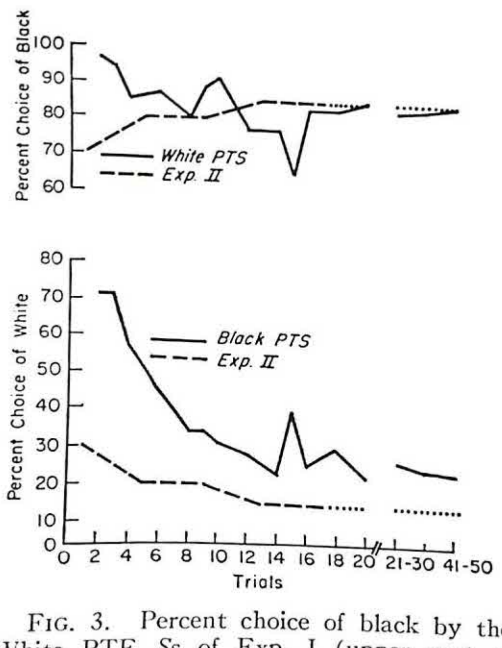


FIG. 3. Percent choice of black by the White PTE *S*s of Exp. I (upper part of figure) and percent choice of white by the Black PTE *S*s (lower part of figure) compared with initial choice of Exp. II *S*s over trials. (The abscissa is plotted in total trials—both test and PTE trials. The dotted lines are hypothetical extensions of the Exp. II *S*s' choices.)

DISCUSSION

The findings of previous studies that *S*s initially alternate at a high level was well substantiated by the present study. A binomial probability test for all groups in Exp. I, with the exception of the forced groups with black PTE, showed that first day alternation was significantly above 50% ($p < .01$). Alternation, however, declined over days, leveling off for the combined brightness groups at about the 50% level. Similarly, in Exp. II, alternation started at 55% and rapidly dropped to 25%.

A major problem in determining whether or not alternation habituates is the designation of a criterion which represents the level of choice *S* would make in the situation without the effects of recent prior exposure. This criterion will be referred to as the Nonsatiation Performance (NSP) level. Most brightness alternation studies, including the present one, have used 50% choice of black and white as the NSP criterion. However, the results of Exp. II suggest that *S*s' empirically determined preferences for black and white should be considered in specifying the NSP base level. Only one known study (Dember & Fowler, 1959) reported a black preference (65%). In Exp. II of the present study the choice behavior on the initial trial of each day suggests that *S*s' preference for the brightnesses used in this study were 80% for black and 20% for white. Consequently, the results of Exp. I were reevaluated using the 80% black–20% white preference levels as NSP criteria.

The upper and lower parts of Fig. 3 present these comparisons. The curves for percent initial trial black-white choice of Exp. II *Ss* represent the NSP levels. On the early test trials alternation (choice of black) for White PTE *Ss* in Exp. I is significantly above 80% ($p < .01$) and alternation (choice of white) for Black PTE *Ss* is significantly above 20% ($p < .001$) by a binomial probability test. Both curves, then, approach and appear to approximate their respective NSP levels over the remaining trials. The White PTE curve was not different significantly from Trial 3 on ($p > .05$) and the Black PTE curve from Trial 12 on. Therefore, when brightness preference is considered in estimating the Nonsatiation Performance level, there is good evidence that alternation is initially present and then habituates over trials.

In Exp. I neither the number of PTE trials nor the method of PTE was shown to be an effective variable. The latter factor was varied in order to investigate the finding that there is more alternation following forced trials than following free trials (Dember & Fowler, 1959; Thompson, 1962). The failure of this variable to produce a difference in percent alternation indicates that the element of choice itself is not a factor affecting alternation behavior.

In conclusion, regardless of which level, 50% or the 80-20% NSP level, is used as a criterion, alternation on the early trials in the present study was found to be significantly above "chance" performance. Here it is possible to appeal to any of several hypotheses which have been offered to explain alternation: i.e., that *S* initially has an exploratory drive, curiosity motivation (Montgomery, 1951), or stimulus satiation (Glanzer, 1953) which results in alternation behavior. The demonstration in both Exp. I and II that alternation does habituate reduces the importance of this behavior in the schema of learning theory. However, the importance of alternation as a behavioral phenomenon in its own right and its possible antagonistic effects on the early trials of learning deserve further study.

REFERENCES

DEMBER, W. N., & FOWLER, H. Spontaneous alternation after free and forced trials. *Canad. J. Psychol.*, 1959, **13**, 151–154.

ESTES, W. K., & SCHOEFFLER, M. S. Analysis of variables influencing alternation after forced trials. *J. comp. physiol. Psychol.*, 1955, **48**, 357–362.

FOWLER, H., BLOND, J., & DEMBER, W. N. Alternation behavior and learning. The influence of reinforcement magnitude, number and contingency. *J. comp. physiol. Psychol.*, 1959, **52**, 609–614.

GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.

MONTGOMERY, K. C. The relation between exploratory behavior and spontaneous alternation in the white rat. *J. comp. physiol. Psychol.*, 1951, **44**, 582–589.

MONTGOMERY, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45**, 287–293.

ROTHKOPF, E. Z., & ZEAMAN, D. Some stimulus controls of alternation behavior. *J. Psychol.*, 1952, **34**, 235–255.

THOMPSON, M. E. Stimulus alternation, response repetition and response alternation in a multiple-choice situation. *Psychol. Rep.*, 1962, **11**, 523–527.

WALKER, E. L. The duration and course of the reaction decrement and the influence of reward. *J. comp. physiol. Psychol.*, 1956, **49**, 167–176.

(Received March 23, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 618–623

CHILDREN'S PERFORMANCE AS A FUNCTION OF THE DEGREE OF VISUAL STIMULUS DEPRIVATION AND SATIATION [1]

RICHARD D. ODOM

University of Minnesota

112 boys and girls (8 through 9 yr.) participated in a 3-stage, lever-pressing task with visual stimuli as reinforcers. The preexperimental period consisted of a response period with no reinforcers followed by a response period with reinforcers. During the experimental period Ss made no lever-pressing responses, and 4 conditions involving deprivation (reinforcers withheld) or satiation (reinforcers presented) for either 2 or 10 min. were introduced: D10, D2, S2, S10. The postexperimental period consisted of a response period with reinforcers. It was found that rate of lever pressing increased as degree of satiation decreased and degree of deprivation increased. The results were interpreted as lending support to the notion that visual stimulus deprivation and satiation can act as a source of motivation.

The concepts of deprivation and satiation have frequently been used to account for behavior that cannot be explained wholly in terms of traditional learning variables. Such operations involving primary reinforcers have been employed in research using rats as *S*s in investigations of the effects of different levels of deprivation upon the acquisition of an instrumental response and upon the asymptotic level of performance (e.g., Kimble, 1951; Ramond, 1954). A number of studies (Erickson, 1962; Gewirtz & Baer, 1958a, 1958b; Hartup, 1958) have investigated the effects of deprivation and satiation of social stimuli on children's subsequent performance in tasks where verbal approval was the reinforcer. Using infrahuman organisms as *S*s, Butler (1957a, 1957b), Kish and Baron (1962), and Fox (1962) found that effectiveness of certain classes of simple visual and auditory reinforcers can be modified with deprivation and satiation operations. In general, the results of the above studies have demonstrated that reinforcers, whether primary, social, or sensory in nature, were more effective following a longer period of deprivation than they were following a shorter period of deprivation or a period of satiation.

More recently, the results of several studies (Horowitz, 1962; Stevenson & Knights, 1961; Stevenson & Odom, 1961, in press) have indicated that different classes of visual stimuli serve effectively as reinforcers for children's behavior. Using children as *S*s, Odom (1964) found that deprivation and satiation operations with visual and auditory stimuli influenced the effectiveness of these stimuli as reinforcers in a subsequent task.

If deprivation and satiation are considered as points on a dimension of reinforcer availability, then the degree of deprivation and satiation can be varied by increasing or decreasing the amount of time in which a potential reinforcer is either unavailable or

[1] This study was supported in part by a grant from the National Institute of Mental Health (M-03519) United States Public Health Service. The author is grateful to the staffs of Roosevelt and Ericsson Elementary Schools, St. Paul, Minnesota, for their cooperation in securing *S*s for this study.

available and responded to by the organism. The present study was designed to test the effects of degree of visual stimulus deprivation and satiation on children's performance in a simple lever-pressing task. For each *S* there were three stages in the experiment: (*a*) a preexperimental period consisting of a response period with no reinforcers followed by a response period with reinforcers, (*b*) an experimental period (either 2 or 10 min.), during which reinforcers were either withheld (deprivation) or presented (satiation), and (*c*) a postexperimental period with reinforcers. Sex of *S* was included as a variable since earlier studies (Stevenson & Odom, 1961, in press) have indicated that visual reinforcers are more effective for boys than for girls.

On the basis of the research presented above, certain predictions were made concerning the results of the present study. Considering both deprivation and satiation as part of a single dimension, it was predicted that the rate of responding would decrease as a function of decreasing degrees of deprivation and increasing degrees of satiation. In addition, it was predicted that visual stimuli would be more effective as reinforcers for boys than for girls.

Method

Subjects.—The *S*s were 56 boys and 56 girls from the third and fourth grades of Roosevelt and Ericsson Elementary Schools in St. Paul, Minnesota. An equal number of boys and girls were randomly selected from the two grade levels.

Apparatus.—The study took place in a $7\frac{1}{2} \times 8$ ft. room of a mobile laboratory belonging to the Institute of Child Development. The apparatus has been described fully in earlier studies (Odom, 1964; Stevenson & Odom, 1961). Generally, it consisted of a wooden frame upon which a vertical and a horizontal panel were mounted. The vertical panel contained a ground glass screen on which stimuli could be projected. One end of a black, open-ended tube, $11\frac{1}{2}$ in. in length, was attached to the vertical panel so that it encircled the glass screen. Extending from the front opening of a toy space helmet was a second, similar tube which was designed to fit inside the first tube. Each *S* was required to wear the helmet in order to reduce the availability of visual stimuli present in the experimental room.

Centered on the horizontal panel, which sloped downward from the bottom of the vertical panel, was a vertical wooden lever. The apparatus was mounted on a large panel which was constructed to fill an opening between the experimental room and an adjoining observation room. A chair for *S* was placed in front of the apparatus, and the relationship between the chair and the apparatus was similar to that of a child's chair and desk in a classroom.

Located in the observation room was a Dunning Animatic filmstrip projector which was used to project colors onto the screen. The filmstrip contained three random sequences of 10 different colors. Between each colored frame of the filmstrip was a blank frame which projected a uniform gray onto the screen. Also located in the observation room was the equipment used to program the presentation of stimuli and to record *S*'s responses.

Procedure.—Equal numbers of boys and girls from each grade level were randomly assigned to four groups. All *S*s responded by pressing the lever during a 4-min. preexperimental period and a 5-min. postexperimental period. The first 2 min. of the preexperimental period served as a base-rate period during which no colors were presented. During the last 2 min. of this period and throughout the postexperimental period colors were presented as reinforcers according to a variable ratio schedule with a mean of one color for every eight responses.

The *S*s were instructed not to press the lever during the experimental period which lasted either 2 or 10 min., depending on the condition to which *S* had been assigned. For *S*s in the two satiation conditions, S10 and S2, colors were presented for 10 and 2 min., respectively, according to a variable interval schedule. These two groups of *S*s received a color on the average of every 6 sec. Throughout the three stages of the study each color was presented on the screen for a duration of 1 sec. The *S*s in the D10 and D2 conditions received no colors for 10 and 2 min., respectively.

Each S came individually to the experimental room. The E began by explaining that when the handle was moved during the game, different colors would appear on the screen. The S was told that he could move the handle as quickly or as slowly as he wished and that he could change hands in case the one moving the handle got tired. The E then presented the helmet, put it over S's head, placed the helmet tube inside the apparatus tube, and pushed S's chair forward so that the lever was easily accessible and so that the helmet tube could not be removed from the other tube. After S was told that the game was ready and to begin moving the handle, E went into the observation room, closing the door behind him.

At the end of the preexperimental period E interrupted S and told him that something had happened to the game and that while it was being fixed, S was not to move or touch the handle. The E then returned to the observation room. At the end of the experimental period E told S the game was ready to play once again, and S then began moving the handle. The E remained in the observation room until the task was over. For each S the number of lever-pressing responses was recorded separately for each minute of the pre- and postexperimental periods.

Results

Preexperimental period.—The number of responses made during the first 2-min., base-rate period of the task provides a measure of performance independent of the effects of visual reinforcers. The results of a 4 (Condition) $\times$ 2 (Sex) analysis of variance of these data showed a significant Sex effect, F (1, 104) = 4.49, $p < .05$, which reflects the difference between the initial rate of lever pressing of boys (M = 110.71) and that of girls (M = 85.68). Neither the Condition effect nor Condition $\times$ Sex effect was significant (Fs < 1.00).

During the 2 min. following the base-rate period, visual reinforcers were made contingent on the lever-pressing response. A 4 (Condition) $\times$ 2 (Sex) analysis of variance was performed on the number of responses made during this period to determine if differences existed between the performance of the groups. The significant Sex effect, F (1, 104) = 7.56, $p < .01$, indicated that the boy's rate of response (M = 132.07) was higher than that of the girls (M = 96.84). None of the other effects was significant (Fs < 1.00).

To determine whether there was an increase in rate of responding during the 2-min. period with reinforcers as compared to the base-rate period with no reinforcers, difference scores were obtained by subtracting the number of responses made during the base-rate period from the number made during the subsequent 2 min. That all groups had higher rates of response during the reinforcement period than during the base-rate period was evidenced by the lack of any negative difference scores. An analysis of variance of these scores yielded a significant Grand Mean effect, F (1, 104) = 51.31, $p < .001$, indicating that rate of response was significantly greater during the period with reinforcers than during the base-rate period. In addition, the Sex effect was once again significant, F (1, 104) = 5.04, $p < .05$. The difference scores were larger for boys (M = 21.36) than for girls (M = 11.16), indicating that boys showed a greater increment in response during the reinforcement period than girls. As was the case in the other two analyses, the Condition and interaction effects were not significant (Fs < 1.00). To determine whether the difference scores were related to the initial base-rate scores, correlations between these scores were performed separately for boys and girls. Neither the correlation coefficient for boys (r = −.05) nor that for girls (r = .15) was statistically significant, indicating that the increment in response made during the reinforcement period was independent of

performance during the base-rate period.

Postexperimental period.—The score used in the analysis of the results of the postexperimental period was a difference score. This score was obtained by subtracting the mean number of responses made during each of the last 2 min. of the preexperimental period from the number of responses made during each minute of the postexperimental period. The mean number of responses made during the last 2 min. of the preexperimental period by *S*s in the four conditions were as follows: D10, 59.27; D2, 53.54; S2, 56.25; S10, 59.86. Each of these values was obtained by averaging the mean number of responses made during each of the 2 min. The mean number of responses made during the entire postexperimental period by *S*s in the four conditions were as follows: D10, 340.30; D2, 300.87; S2, 287.13; S10, 274.85.

The difference score represents an increase or decrease in rate of responding during the postexperimental period as compared to rate of responding during the reinforcement period of the preexperimental period. It provides a measure of the influence that the experimental conditions had on the rate of responding.

TABLE 1

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES FOR POSTEXPERIMENTAL PERIOD

Source	*df*	*MS*	*F*
Condition (C)	3	5243.93	4.59**
Dep. 10 vs. Dep. 2	1	325.73	<1.00
Sat. 10 vs. Sat. 2	1	2580.36	2.26
Dep. 10+2 vs. Sat. 10+2	1	12825.72	11.23**
Sex (S)	1	1440.01	1.26
C×S	3	1952.04	1.70
Error	104	1141.64	
Minutes (M)	4	4163.39	16.82***
C×M	12	90.04	<1.00
S×M	4	179.63	<1.00
C×S×M	12	104.22	<1.00
Error	416	247.51	
Total	559		

** $p < .01$.
*** $p < .001$.

TABLE 2

MEAN DIFFERENCE SCORES FOR THE FOUR GROUPS ACROSS THE 5 MIN.

Groups	Min. 1	2	3	4	5
D 10	−2.17	8.60	8.42	12.64	16.46
D 2	−.21	6.39	8.50	7.82	10.67
S 2	−10.82	.57	4.50	5.10	6.53
S 10	−14.92	−7.10	−2.57	−2.28	2.42

The mean difference scores for *S*s in the four experimental groups were as follows: D10, 43.95; D2, 33.17; S2, 5.88; S10, −24.45. By inspecting the relationship between these means it can be seen that rate of responding decreased as deprivation decreased and as satiation increased. These data were analyzed by means of a 4 (Condition) × 2 (Sex) × 5 (Minutes) analysis of variance employing orthogonal comparisons. Comparisons were made between the performance of the D10 and D2 groups, the S10 and S2 groups, and the combined D10, D2 and combined S10, and S2 groups. The results of the analysis are summarized in Table 1.

It can be seen in Table 1 that the Condition effect was significant, indicating that the groups differed in their overall rate of response. The significant D10, D2 vs. S10, S2 comparison was based on the fact that the total of the combined difference scores of the deprivation groups was greater than that of the satiation groups.

The only other significant effect was the Minutes effect. It can be seen in Table 2, which presents the mean difference scores for the four groups across the 5 min., that all groups showed increasing increments of response throughout the postexperimental period.

Discussion

The analysis of the base-rate scores indicated that the boys had a significantly higher rate of response than the girls. A possible explanation for this finding is that boys found the task more interesting than the girls due perhaps to parts of the apparatus that could be associated with space equipment.

When rate of response for the preexperimental period with reinforcers was compared with rate for the base-rate period, the difference scores were higher for boys than for girls; indicating that the introduction of visual reinforcers resulted in greater increments in response for boys than for girls. This finding lends support to the reliability of the results of earlier studies (Stevenson & Odom, 1961, in press) which showed visual stimuli to be more effective as reinforcers for boys than for girls.

The lack of significant Condition and Condition × Sex effects in the three analyses of the data for the preexperimental period was anticipated, since conditions were identical for all groups during this period, and there was therefore, no reason to expect differences in performance among the four groups.

That the various conditions affected the rate of incremental responding, or, in other words, the effectiveness of the reinforcers, was demonstrated by the differences in rate of responding among the four groups. That the four groups demonstrated an increase in increment of response across minutes is assumed to be due, at least partially, to the ability of the visual reinforcers to maintain and strengthen the response regardless of the condition.

In addition to the present study, several studies, cited earlier, have indicated that visual stimuli have reinforcing characteristics that are effective in modifying children's behavior. In the present study *S*s in the four groups showed higher rates of responding during the preexperimental period when visual reinforcers were presented than during the base-rate period when no reinforcers were presented. Behavior was further modified by the deprivation and satiation operations employed during the experimental period, so that rate of responding increased as a function of decreasing satiation and increasing deprivation.

In light of these results it seems reasonable to speculate about the possibility of visual stimulus deprivation and satiation as a source of motivation. Throughout the task the amount of visual stimulation was reduced considerably from the amount normally available to *S*s, with the greatest reduction occurring for the D10 group and the least reduction occurring for the S10 group. It is possible that certain processes within the organism may adapt to a particular level of visual stimulation. When the amount of stimulation falls below this level, the organism may become generally responsive, with certain responses being reinforced by visual stimuli which consequently increase the probability of the reoccurrence of that response. On the other hand, when such a need state exists, it may activate specific responses that are designed to obtain specific stimuli.

The results of the present study indicate that visual deprivation and satiation have certain behavioral effects that are much like those of similar operations using other classes of reinforcers. It is possible that a process similar to that occurring with deprivation and satiation operations using primary reinforcers also exists when sensory reinforcers are used.

References

Butler, R. A. Discrimination learning by rhesus monkeys to auditory incentives. *J. comp. physiol. Psychol.*, 1957, **50**, 239–241. (a)

Butler, R. A. The effect of deprivation of visual incentives on visual exploration motivation in monkeys. *J. comp. physiol. Psychol.*, 1957, **50**, 177–179. (b)

Erickson, M. T. Effects of social deprivation and satiation on verbal conditioning in children. *J. comp. physiol. Psychol.*, 1962, **55**, 953–957.

Fox, S. S. Self-maintained sensory input and sensory deprivation in monkeys: A behavioral and neuropharmacological study. *J. comp. physiol. Psychol.*, 1962, **55**, 438–444.

GEWIRTZ, J. L., & BAER, D. M. Deprivation and satiation of social reinforcers as drive conditions. *J. abnorm. soc. Psychol.*, 1958, **57**, 165–172. (a)

GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, **56**, 49–56. (b)

HARTUP, W. W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. *Child Develpm.*, 1958, **29**, 191–201.

HOROWITZ, F. D. Incentive value of social stimuli for preschool children. *Child Develpm.*, 1962, **33**, 111–116.

KIMBLE, G. A. Behavior strength as a function of the intensity of the hunger drive. *J. exp. Psychol.*, 1951, **41**, 341–348.

KISH, G. B., & BARON, A. Satiation of sensory reinforcement. *J. comp. physiol. Psychol.*, 1962, **55**, 1007–1010.

ODOM, R. D. Effects of auditory and visual stimulus deprivation and satiation on children's performance in an operant task. *J. exp. child Psychol.*, 1964, **1**, 16–25.

RAMOND, C. K. Performance in selective learning as a function of hunger. *J. exp. Psychol.*, 1954, **48**, 265–270.

STEVENSON, H. W., & KNIGHTS, R. M. Effect of visual reinforcement on the performance of normal and retarded children. *Percept. mot. Skills*, 1961, **13**, 119–126.

STEVENSON, H. W., & ODOM, R. D. Effects of pretraining on the reinforcing value of visual stimuli. *Child Develpm.*, 1961, **32**, 739–744.

STEVENSON, H. W., & ODOM, R. D. Visual reinforcement with children. *J. exp. child Psychol.*, in press.

(Received March 24, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 624–629

CONCEPT IDENTIFICATION: THE EFFECTS OF VARYING LENGTH AND INFORMATIONAL COMPONENTS OF THE INTERTRIAL INTERVAL [1]

LYLE E. BOURNE, Jr., DONALD E. GUY

University of Colorado

DAVID H. DODD

University of Missouri, Kansas City

AND DON R. JUSTESEN [2]

Veterans Administration Hospital, Kansas City, Missouri

In Experiment I, Ss solved concept-identification problems under conditions formed by combining: 4 intertrial durations (1–25 sec.); 2 levels of task complexity (1 and 5 irrelevant stimulus dimensions); and 2 modes of controlling duration of stimulus patterns (self-paced vs. fixed interval). Performance (a) improved, then got worse with increases in the interval, the optimal length being greater in more complex problems, and (b) was unaffected by mode of stimulus control. In Experiment II, Ss served in 4 intertrial conditions: (a) simple time-out alone; (b) display of stimulus pattern; (c) display of signal indicating response correctness; or (d) both b and c. 3 intertrial durations were used for each condition: 1, 15, or 29 sec. Trends were the same, but performance did not worsen during longer intervals under Conditions b and d.

A single trial in a typical concept-identification problem includes, in the order given, the following critical events: (*a*) the appearance of a stimulus pattern; (*b*) *S*'s categorizing response to it; and (*c*) the immediate withdrawal of the pattern and presentation of informative feedback (IF) that signals the correct response. Then a brief intertrial interval, often referred to as the postfeedback interval, intervenes before commencement of a new trial.

Bourne and Bunderson (1963) found that increasing the length of the intertrial intervals significantly reduced number of errors and trials to solution in identification problems. The effect, moreover, was related to task complexity, since greater facilitation was observed in problems complicated by larger numbers of irrelevant stimulus dimensions, i.e., variations among stimulus patterns not related to problem solution.

The restricted range of intertrial

[1] Part of the research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication 43 of the Institute. This research was supported in part by Grants MH-01759-06 and MH-08315-01 from the National Institute of Mental Health, United States Public Health Service; by funds provided by the Committee on Research, University of California; and by 8,200 research funds provided by the United States Veterans Administration.

The authors express appreciation to S. Goldstein, who collected the data for Exp. I, and to F. McEvoy, who assisted in Exp. II. The *S*s serving in Exp. I were students at the University of California, Berkeley. The authors are grateful for the assistance and facilities provided by the Department of Psychology. Institutions in the Missouri-Kansas area supplying *S*s for Exp. II were the University of Kansas, Park College, and the University of Missouri at Kansas City. The authors wish to express their gratitude to the several cooperating administrations.

[2] Also at Kansas University Medical Center, Kansas City.

durations explored in that study, 1–9 sec., was limited by a requirement that conditions overlap those of an earlier experiment (Bourne, 1957). Thus available information on the general form of the relationship between performance and the intertrial interval (and on their interaction with task complexity) is incomplete. Obviously, the facilitation reported by Bourne and Bunderson must diminish beyond 9 sec., if only because of a ceiling effect. A more compelling reason to explore a greater range of times in concept-identification problems is provided by the possibility that longer intertrial intervals may interfere with performance, an effect which has been observed in other learning tasks (e.g., Bilodeau & Bilodeau, 1958). There is, moreover, the question of the composition of the intertrial interval—the inclusion or exclusion of various, task-related sources of information may greatly affect performance. It was the purpose of these experiments to explore further the role of intertrial variables in concept identification.

General Method

Task and apparatus.—A complete description of procedural details has been given (Bourne, 1957). Briefly, a 16-mm. stripfilm projector was used to present *S* with a series of stimulus patterns (geometric designs) each of which was a combination of the levels of two relevant plus one or more irrelevant binary stimulus dimensions, such as color (red-green) and form (square-triangle). To identify the category to which any particular pattern belonged, *S* pressed one of four, unlabeled response keys; the task was to discover the two relevant dimensions and how the four conjunctive combinations of their levels were associated with the four keys. Following each response, *S* was immediately informed of the correct category for the stimulus pattern by the onset of an IF light over an appropriate key. The sequence of IF signals was controlled by a Western Union tape transmitter equipped with a precoded tape that matched the stimulus series. As in earlier studies, both levels of irrelevant dimensions appeared within the sequence of stimulus patterns but were uncorrelated with IF signals. Electronic timing units controlled duration of IF (1 sec. in Exp. I and in control conditions of Exp. II), and the length of the intertrial interval. The *S*'s responses were recorded on an Esterline-Angus operations recorder. All *S*s were initially given detailed instructions about the task, the meaning of IF signals, and the criterion of problem solution—16 consecutively correct responses.

Experiment I

The major aim of Exp. I was to explore, in concept-identification problems characterized by different degrees of complexity, a more extensive range of intertrial interval than used in previous studies. It had the secondary purpose of gathering evidence on the relationship between *S*'s activities during this interval, and another, the time from onset to termination of the stimulus pattern. Usually, the length of stimulus exposure has been determined by *S*, terminating only when *S* made a response. The effect of varying the intertrial interval, no matter how interpreted, must be somewhat independent of stimulus duration, for *S*s apparently cannot or do not compensate for short intertrial durations by lengthening the stimulus interval (Bourne & Bunderson, 1963). It may be, however, that fixing the stimulus interval at some brief length would interfere with performance when the intertrial interval is also brief, thus implying an interaction between *S*'s activities during the two periods. To assess this possibility, the plan of the present experiment included parallel conditions with *E*-determined (fixed) and *S*-determined (self-paced) stimulus intervals.

Method

Subjects.—The 192 *S*s participating in the experiment were volunteers from introductory

psychology courses. Each was assigned to one of 16 experimental conditions and served individually for one session.

Design.—A $4 \times 2 \times 2$ factorial design combined four lengths of intertrial interval (1, 9, 17, or 25 sec.), two degrees of task complexity (problems with one or five irrelevant stimulus dimensions), and two stimulus interval conditions (self-paced, or fixed at 5 sec. on all trials). The *S*s performing on the fixed interval were instructed to respond while the stimulus was on the viewing screen and that failure to respond would be counted as an error; fewer than 1% of observed errors involved failure to respond. The 5-sec. fixed duration was selected because an informal pilot study revealed this to be the approximate mean length of *S*-determined intervals on precriterion trials. Lengths of the intertrial interval were chosen after an inspection of data from the same pilot study, which suggested the existence of an optimal duration of less than 25 sec. for problems with five irrelevant dimensions.

Two problems were used for each level of complexity; the two were equivalent save for differing relevant dimensions. Each was used equally often in all 16 main conditions—to reduce the possibility of collusion among *S*s

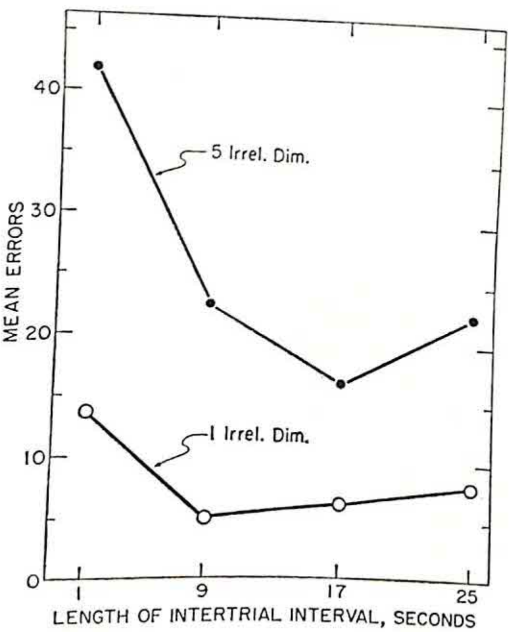


FIG. 1. Mean number of errors to solution as a joint function of length of intertrial interval and number of irrelevant stimulus dimensions. (Each plotted point represents the data of 24 *S*s.)

and to increase the generality of conclusions warranted by the data.

Results

Errors to solution were reliably affected by three factors: number of irrelevant dimensions, $F\,(1, 160) = 151.89$; length of the intertrial interval, $F\,(3, 160) = 54.16$; and the interaction of these two variables, $F\,(3, 160) = 10.13$; all with $p < .01$. The interaction of irrelevant dimension and intertrial interval was further analyzed into orthogonal components of trend: both the linear and quadratic terms were reliable; $F_{\text{Irrel.} \times \text{Intertrial:Lin.}}\,(1, 160) = 18.53$ and $F_{\text{Irrel.} \times \text{Intertrial:Quad.}}\,(1, 160) = 10.08$, $p < .01$. The influence of these main variables is summarized in Fig. 1. The function relating errors to length of the intertrial interval has a minimum, the value of which increases as does the overall difficulty of the task, with number of irrelevant dimensions.

No other main effect or interaction contributed significant variance to performance data as indexed by errors. Analysis of trials-to-solution data yielded the same outcome.

Discussion

The results of Exp. I essentially replicated an earlier finding (Bourne & Bunderson, 1963) that a moderate increase in the length of the intertrial interval (to 9 sec.) produced reliable improvement in overall performance on concept-identification problems. Further, the data indicate the lack of a substantial interaction between *S*'s activities during intertrial and stimulus intervals, for the performance trends were the same when duration of the stimulus was fixed or self-determined. As mentioned by Bourne and Bunderson, there are several plausible interpretations of facilitation due to lengthened intertrial intervals; for example, the interval may provide opportunity: to associate characteristics of the stimulus with the

signaled response category; to rehearse retained information; and/or to formulate new hypotheses about solution. These possibilities, however, are incomplete, since they predict only an asymptote and not the deleterious effect observed when intertrial intervals exceed 9–17 sec. A second controlling process apparently determines the optimizing of performance, but available evidence fails to specify its source(s).

Experiment II

In Exp. II, the intertrial interval not only was varied in length, but also incorporated the four possible conditions involving presence or absence of stimulus pattern and corresponding IF signal from the preceding trial. The purpose was to determine whether availability of either source of information would counteract the interference effect associated with longer intertrial intervals. Such a result would provide indirect evidence on the importance of memory for previously given information as a determiner of rate of problem solving.

Method

Subjects.—The *S*s were 144 undergraduate college students who volunteered for participation in the experiment. Seventy-two *S*s were drawn from elementary psychology courses offered at the University of Colorado and 72 were enlisted from various colleges and universities in the Missouri-Kansas area. Each *S* was assigned randomly to one of the 12 main conditions, and served individually for a single session.

Design.—The experimental plan combined factorially three lengths of intertrial interval (1, 15, and 29 sec.) with the presence or absence (during the intertrial interval) of either or both the stimulus pattern and the IF signal. Thus, the conventional condition for concept-identification experiments, where neither pattern nor IF signal is available during the intertrial interval, can be compared systematically with conditions wherein either or both remain on display. The lengths of intertrial intervals were selected to overlap the range of times used in Exp. I. Two

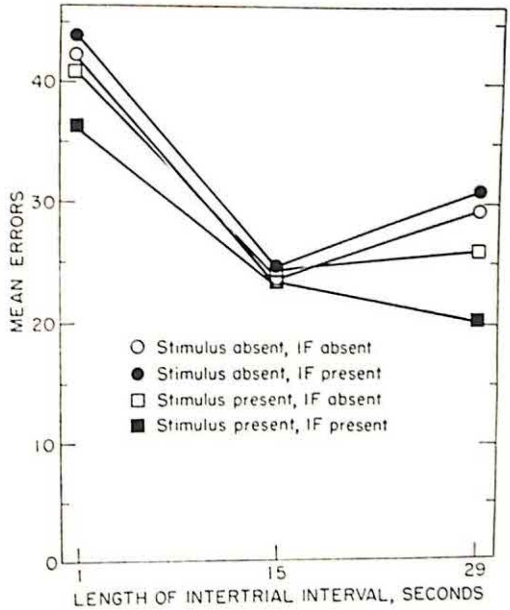


FIG. 2. Mean errors to solution as a function of length of intertrial interval and the presence or absence of stimulus pattern and/or IF signal on any trial. (Each plotted point represents the data from 12 *S*s. All conditions utilized five irrelevant stimulus dimensions.)

equivalent problems (differing only in terms of the dimensions chosen to be relevant but with five irrelevant dimensions in each case), were used equally often for *S*s in all main conditions. The source of *S*s was systematically balanced, with equal *N*s from the two locations serving in each condition.

Results

Analysis of variance on errors to solution revealed the effect of length of intertrial interval, $F(2, 96) = 30.78$, $p < .01$, to be essentially the same as that observed in Exp. I. Overall performance was best with the intermediate (15-sec.), and worst with the shortest (1-sec.) interval. There was, in addition, a general facilitative effect due to the availability of the stimulus pattern during the intertrial interval, $F(1, 96) = 4.35$, $p < .05$. The presence or absence of the accompanying IF signal was an ineffective variable. There were no reliable

interactions between the three main variables. However, the conditions wherein the stimulus pattern did or did not persist during a 29-sec. intertrial interval did differ reliably, $F(1, 96) = 3.95$, $p < .05$. Figure 2 portrays errors to solution as a function of the three main variables; performance did not degrade during the longest intertrial interval when the stimulus pattern of the preceding trial was continuously available.

Performance on the two problems, in spite of their formal equivalence, was reliably different in the present study, $F(1, 96) = 9.87$, $p < .01$, and there is also a significant interaction between problems and the location from which Ss were recruited, $F(1, 96) = 4.29$, $p < .05$. Further analysis of this interaction demonstrated that the problem difference was considerably larger in the Kansas-Missouri than in the Colorado sample.

Findings in an analysis of variance based on total trials to criterion were the same as those based on the error analysis, except that the presence or absence of the stimulus pattern during the intertrial interval failed to produce a reliable difference.

Discussion

The data of Exp. I and II clarify some of the relations between intertrial variables and concept-identification performance. Under conventional procedure (where neither stimulus pattern nor feedback signal is availed *S* during the intertrial interval), performance improved and then worsened as the intertrial interval lengthened from 1 through 29 sec. In Exp. I, the optimum of this curvilinear relation was found to shift as a function of problem complexity; on the easier of two tasks, optimal performance was associated with a much shorter interval.

Improvement in performance produced by moderate increases in intertrial duration very likely results from enhanced opportunity for *S* to associate the stimulus pattern with the signaled response category, and/or make appropriate inferences about relevant or irrelevant aspects of the stimulus. The observed optimizing and subsequent deterioration of performance under longer intertrial intervals strongly suggests the influence of a second controlling process—an interference effect that accumulates during and across intertrial intervals, eventually overcoming the gains wrought by moderate intertrial durations. In simpler problems, with fewer irrelevant dimensions and less variable stimulus patterns, less benefit would derive from intertrial practice; interference effects, therefore, would more likely be manifest at shorter intertrial durations.

While it is difficult to specify the source of such interference effects, loss of memory (for information provided by previously displayed pattern stimuli) is quite likely. The results of Exp. II demonstrated that performance did not deteriorate, even with the longest intertrial interval used, when stimulus patterns remained available to *S*. Indeed, fewest mean errors occurred in the condition where both stimulus and IF signal were on display throughout at 29 sec. intertrial interval. However, performance differences accruing to presence or absence of IF signals per se were statistically negligible.

The data are open to several interpretations. It may be that Ss deprived of stimuli during intertrial intervals retain but a fraction of total information availed in previous intratrial displays. If forgotten aspects of stimuli are relevant, or if remembered aspects are irrelevant, and if either is improperly associated with a response category, there would certainly result a delay in problem solution. Such an interpretation is consistent with related findings of Cahill and Hovland (1960), and of Bourne, Goldstein, and Link (1964), which demonstrated that the majority of errors made in concept learning are due to utilization of hypotheses that are incompatible with information provided by previously seen but no longer available

stimuli. Incorrect hypotheses, thus, are largely attributable to S's failure to retain or retrieve information about earlier displays of pattern stimuli.

Finally, the results of the present study may have been due in part to another controlling factor: loss of attention or vigilance to task. Prolonged intertrial intervals devoid of task-related stimuli might well conduce to such a loss. This and other possible sources of interference remain to be assessed in further experiment.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

BOURNE, L. E., JR., & BUNDERSON, C. V. Effects of delay of informative feedback and length of postfeedback interval on concept identification. *J. exp. Psychol.* 1963, **65**, 1–5.

BOURNE, L. E., JR., GOLDSTEIN, S., & LINK W. E. Concept learning as a function of availability of previously presented information. *J. exp. Psychol.*, 1964, **67**, 439–448.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

(Received March 26, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 630–636

EFFECTS OF ORDERED AND CONSTANT SUCROSE CONCENTRATIONS ON NONREINFORCED PERFORMANCE [1]

TOM N. TOMBAUGH [2] AND MELVIN H. MARX

University of Missouri

2 experiments were performed to determine the effects of different sucrose concentrations on nonreinforced performance. In the 1st experiment, the concentrations (4%, 8%, 32%, 64%) were presented in either an increasing or decreasing order on alternate days—all Ss experienced both orders. In the 2nd experiment, Ss received only 1 of the above concentrations. The results indicated that the response rate preceding extinction was not an adequate basis for predicting nonreinforced performance. An alternative view, which focused on the stimulus properties of the sucrose reinforcers as predictors of resistance to extinction, is presented.

The present studies were designed to determine if nonreinforced performance can be better predicted from the response rate immediately preceding extinction or from the value of the sucrose incentive preceding extinction. In his review article on magnitude of reinforcement, Pubols (1960) concluded, along with Metzger, Cotton, and Lewis (1957), that resistance to extinction is best predicted by the terminal response rate—the rate of responding immediately preceding extinction. Others (Campbell, 1958; Guttman, 1953; Young & Green, 1953) indicate that some dimension of the reinforcement properties—taste, receptor sensitivity, or a motivational property, is a more important variable. Marx, Tombaugh, Cole, and Dougherty (1963) trained two groups of rats to bar press on either an increasing (4%, 8%, 32%, 64%) or decreasing (64%, 32%, 8%, 4%) order of sucrose incentives, with concentrations progressively changed between each of four consecutive 30-sec. nonreinforced test trials. Although *S*s bar pressing on the increasing order had a reliably *lower* terminal response level than those on the decreasing order, they pressed *more* during the nonreinforced trials. These results suggest a motivational determination of resistance to extinction, and support the interpretation that extinction is essentially a function of a motivational decrement (Marx, 1958, 1960). That is, *S*s which received progressively more desirable incentives (increasing order) were more motivated to respond during extinction than *S*s offered progressively less desirable incentives (decreasing order).

The first experiment incorporated the same basic design except that every *S* was exposed to *both* incentive orders. Thus, on alternating days each *S* was given one of the two orders. The purpose of this daily alternation was to determine whether the previously found differences between training and extinction would still vary

[1] This research was supported by United States Public Health Service Grant MH00817-10 from the National Institute of Mental Health, United States Public Health Service Grant HD00895-10 from the National Institute of Child Health and Human Development, and National Science Foundation Grant GB229 from the National Science Foundation. The authors wish to thank Judith E. Peterson for her assistance in collecting the data from Exp. II.

[2] United States Public Health Service Predoctoral Fellow.

systematically as a function of the incentive order, or if experience with both orders would eliminate these differences. That is, we attempted to determine whether the previously demonstrated behavior is a reversible phenomenon, as one would expect if some motivational aspect of the reinforcement were assumed to be the controlling factor. The present design has the important methodological advantage of permitting each *S* to serve as its own control.

Experiment I

Method

Subjects.—The *S*s were 16 naive female albino rats purchased from the Sprague-Dawley Company, which were approximately 90 days old at the beginning of the experiment.

Apparatus.—The apparatus was the same as described by Marx et al. (1963). In brief, it consisted of eight operant-conditioning boxes in which each bar press (BP) mechanically operated a dipper located below the bar. The bar could be made inoperative by means of a mechanical lock. A light above the bar served as a discriminative cue.

Procedure

Training.—Three days prior to the beginning of the experiment *S*s were placed on 23-hr. water deprivation. All *S*s were given 5 days of dipper training. On each day, water-loaded dippers were manually presented 20 times, with the bar absent, at 60-sec. intervals. For the next 5 days *S*s received BP training. For 20 min. they were placed in the boxes with the bar positioned and allowed to bar press continuously for water. Following the termination of BP training, discrimination training began. On the first day of discrimination training, the 20-min. BP session was broken down into eight 120-sec. response trials separated by seven 30-sec. nonresponse intervals. During response trials the bars were operative and the discrimination light was off. During nonresponse trials the bars were inoperative and the light was on. Over the first 4 days the light-off, bar-operative time was gradually decreased from 120 sec. to 30 sec., while the light-on, bar-inoperative time was increased from 30 sec. to 120 sec. On the fifth and sixth days of discrimination training there were eight 30-sec. response trials separated by seven 120-sec. nonresponse trials. Following the sixth and last day of discrimination training *S*s were placed on ad-lib water and 23-hr. food deprivation. Two days later the testing phase was initiated.

Testing.—On the first day half of the *S*s (Group A) received an increasing order of sucrose (4%, 8%, 32%, 64%) on the first four consecutive response trials. The other half (Group B) received a decreasing order (64%, 32%, 8%, 4%). Reinforcement was continuous. No reinforcement was given during the last four response trials. On Day 2, *S*s which had previously received the increasing order, now received the decreasing order, and vice versa. During the 12 test days the orders were alternated daily. The number of BPs during each of the eight response trials was recorded.

Results

Figure 1 shows the mean BPs for each of the incentive orders summed over days. The results were clearly as predicted, confirming the earlier results of Marx et al. (1963). There was a lower terminal reinforced rate for the increasing order than for the decreasing order, but this relationship was reversed in the following nonreinforced trials.

An analysis of variance appropriate to a three-factor, related-measures design was performed on the data for the 12 test days. All *F*s indicated reliable differences except the order variable summed over reinforcement and trials. The *F* of major interest was the Order × Reinforcement interaction. This interaction was reliable, $F\ (1, 15) = 12.13$, $p < .005$, indicating a change in performance from reinforced to nonreinforced trials as a function of incentive order. A breakdown of the analysis, using the appropriate error terms from the major analysis, indicated that the total number of BPs during reinforcement did not differ reliably for the two orders, $F\ (1, 15) = 1.70$, $p > .05$, but the increasing order did produce reliably more BPs than the decreasing order

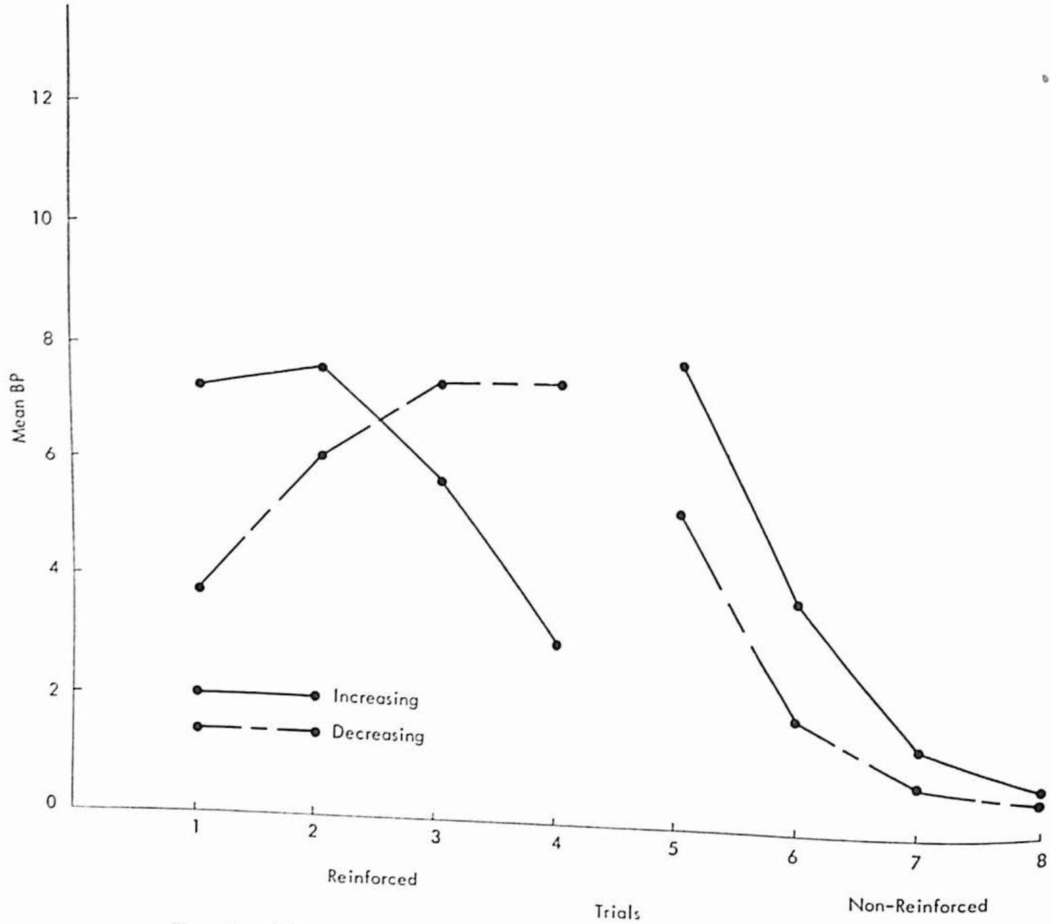


FIG. 1. Mean number of bar presses for 12 days as a function of incentive order and trials.

during nonreinforcement, F (1, 15) = 17.72, $p < .001$. Also, the interaction between order and the last reinforced trial and first nonreinforced trial was reliable, F (1, 15) = 83.61, $p < .001$, indicating an immediate shift in bar-press rate. During the last reinforced trial *S*s pressed less under increasing order than under decreasing order, but they pressed more in the first nonreinforced trial after the increasing order.

The BP scores on the last reinforced and first nonreinforced trial, averaged over all of the increasing and decreasing order sequence separately, were compared for each *S*. These data, in which each *S* served as its own control, offer convincing confirmation of the group differences: for the increasing incentive order, 15 of the 16 *S*s averaged more BPs on the first nonreinforced than on the final reinforced trial, while for the decreasing incentive order the opposite result occurred for 14 of the 16 *S*s. A Q test for related samples (Cochran, 1950) showed these differences to be reliable ($p < .001$).

A similar analysis considering only the first 2 test days revealed that the same relationship between the last reinforced and first nonreinforced trial was present from the beginning of test. On the increasing incentive order 15 of the 16 *S*s bar pressed more on the

first nonreinforced trial than on the last reinforced trial, while on the decreasing order the opposite result occurred for 9 of the 16 *S*s. A *Q* test showed these differences to be reliable ($p < .05$).

It was hypothesized that the results of the first study were primarily attributable to the terminal concentration, independent of ordering, and that the nonreinforced responding was related directly to sucrose concentrations. To test these assumptions the same design was again used, except that no order was incorporated. Four groups of *S*s were used. Each group received one of the following concentrations on all four of the reinforced trials: 4%, 8%, 32%, 64%. It was expected that the decreasing and increasing groups and the constant 4% and 64% groups, respectively, would demonstrate the same function for the terminal reinforced trial and the nonreinforced trials. The 8% and 32% groups were included in order to provide more data, so that the relationship between the terminal concentration and resistance to extinction could be more adequately assessed.

Experiment II

Method

Subjects and apparatus.—The *S*s were 32 naive female albino rats which were approximately 120 days old at the beginning of the experiment and were purchased from the

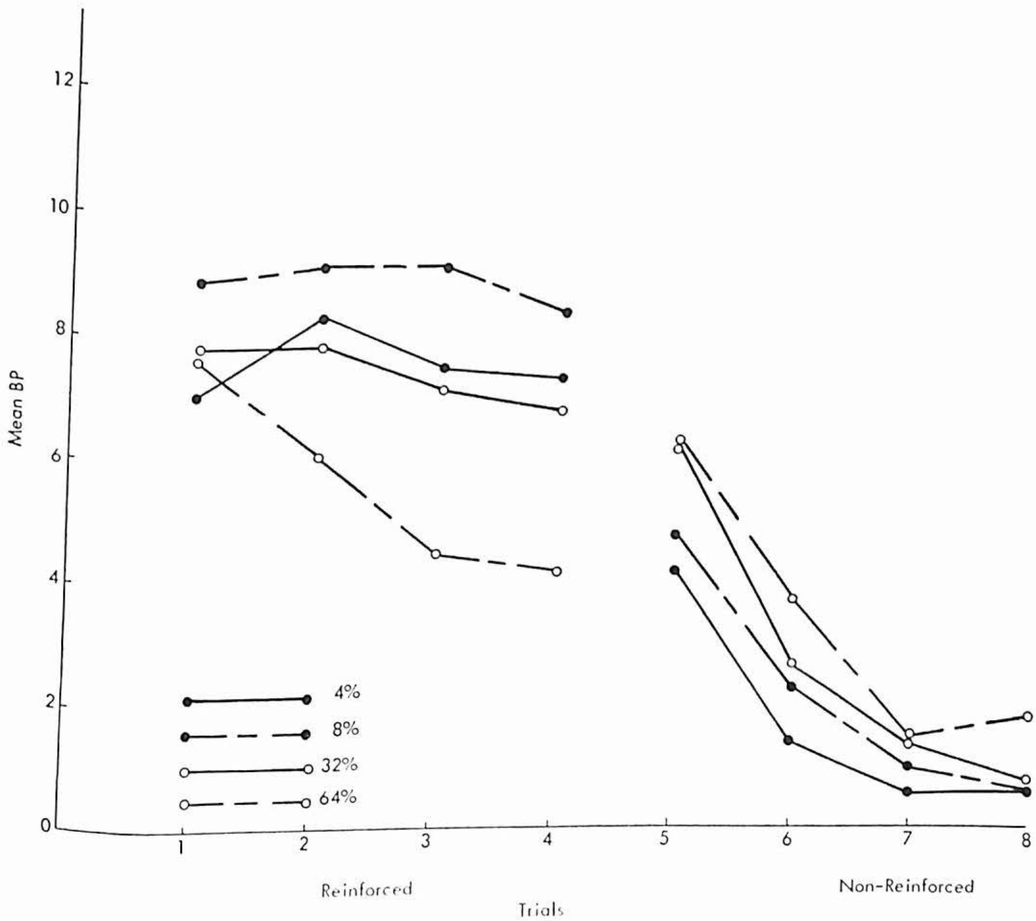


FIG. 2. Mean number of bar presses for 12 days as a function of sucrose concentrations and trials.

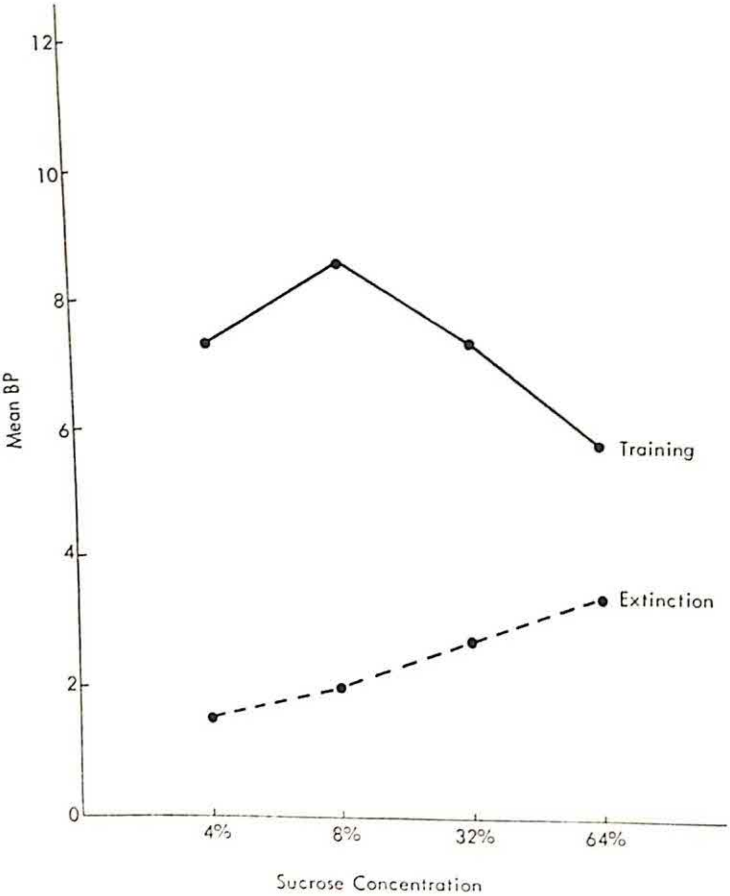


FIG. 3. Mean number of bar presses during training and extinction as a function of sucrose concentrations.

Sprague-Dawley Company. The apparatus was the same as was used in Exp. I.

Procedure

Training and test.—The same training procedure as used in Exp. I was incorporated. In test, the 32 *S*s was divided into four groups designated on the basis of the different sucrose concentrations employed: 4%, 8%, 32%, 64%. On each of the first four consecutive 30-sec. response trials the groups received the *same* sucrose concentrations. The *S*s were not reinforced during the last four consecutive response trials. The *S*s were tested 6 days a week for 4 wk.

Results

The number of BPs per *S* for each response trial was summed over the first 12 days. Group means are shown in Fig. 2. An analysis appropriate to a three-factor experiment in which two factors are repeated measures was performed on these data. The variables under consideration were concentration, reinforcement, and trials. All differences were reliable, but the main interest was the Concentration × Reinforcement interaction, $F(3, 28) = 10.37$, $p < .005$. This analysis reflected the highest rate of reinforced performance by the 8% group, followed by the 4%, the 32%, and the 64% groups, respectively, while during nonreinforcement, BP rate was directly related to concentration. Also, the interaction between trials and concentration was reliable for the reinforced trials, $F(9, 80) = 3.66$, $p < .005$. The interaction was due to

the stable performance over trials by the 4%, 8%, and 32% groups as opposed to a large decrement in responding by the 64% group.

Figure 3 shows the mean performance as a function of sucrose concentration, collapsed over trials. The group differences during reinforcement were reliable, F (3, 28) = 28.69, $p < .005$, as were the differences during nonreinforcement, F (3, 28) = 10.70, $p < .005$. It should be noted that the differences during nonreinforcement were a linear function of the sucrose concentration.

An analysis based on the same information for Wk. 3 and 4 showed a continuation of the reinforced behavior, but no reliable differences during nonreinforcement were present, due to a general lack of responding for all groups.

Discussion

Although Pubols (1960) has indicated that terminal response rate (rate of reinforced responding immediately preceding extinction) predicts resistance to extinction, the present experiments question the generality of such a view. The inadequacy of the response-rate factor is clearly illustrated in both experiments by the existence of an interaction between the last reinforced trial and the first nonreinforced trial. In Exp. I, when all the nonreinforced trials were considered, lower terminal response rates were associated with reliably more bar pressing than higher terminal response rate. This same trend was also present in Exp. II. Thus a lower terminal reinforced response rate may, under certain conditions, be followed by a greater number of nonreinforced BPs.

The present results suggest that the critical factor influencing resistance to extinction is the terminal sucrose concentration used as the incentive. Several facts support this view. In Exp. I, the interaction between the last reinforced and first nonreinforced trials was present even though *Ss* received different incentive orders *on alternating days. This* suggests the pervasive influence of the incentives. It should be noted that the BP rates for each of the various concentrations, irrespective of the order (cf. Points 4%, 8%, and 32% concentration, Fig. 1), were remarkably consistent when considered over the 12 test days. This fact suggests that the incentive shift had relatively little influence upon the BP rate for the different concentrations, and that the terminal concentration (rather than the ordering of the incentives) produced the interaction effects and extinction differences. Experiment II, which was designed to test the above hypothesis, appears to confirm it if one compares the 4% and 64% groups on the last reinforced trial and the nonreinforced trials. These groups have essentially the same relationship as the decreasing and increasing groups in Exp. I and the Marx et al. study. Thus it may be concluded that the terminal level of sucrose is an important, if not the most important, determinant of the interaction and extinction differences.

Figure 3 strikingly demonstrates the relationship between the terminal concentration and resistance to extinction. Here, there exists a direct linear function between the sucrose concentration and the number of nonreinforced BPs. It may be that the sucrose concentration is a good indicator of resistance to extinction, in that the greater the concentration the greater the resistance to extinction, providing the concentrations are discriminable, i.e., logarithmically scaled. However, this is merely speculation on our part, and before any such generalization is made further experimentation with other designs is necessary.

The results of both experiments indicate the necessity of incorporating some type of nonassociative explanation to account for the differential persistence of nonreinforced responding. Any of the several current motivation, frustration, or cognitive theories could be applied in interpreting the data. With respect to the present experiments, all of these theories would assume that the groups

receiving the higher concentrations at the end of training would be more highly motivated (or have greater expectations) than those receiving the lower concentrations. However, greater motivation is not readily apparent in the present experiments, since the number of BPs on the last reinforced period is *inversely* related to concentration, with the possible exception of the 4% group in Exp. II.

The typical explanation for the failure of BP rate to reflect motivation is physiological satiation, which should produce a between-trial BP decrement. Since the only appreciable decrement in response rate is for the 64% group, physiological satiation does not explain all of the results. A second explanation, which applies only to experiments in which continuous reinforcement is used, is that consummatory behavior constitutes an interfering response relative to the BP response. There is evidence that the duration of consummatory activity is a direct function of increased sucrose concentration. Smith and Duffy (1957) demonstrated that with high concentrations of sucrose, *S*s had more persistent drinking behavior than with lower concentrations. Unpublished data from our laboratory support this finding for all solutions of 8% or above. Consequently, the low response rate for high concentrations may be an indirect effect resulting from satiation and perseverative licking behavior. Studies which have minimized the effects of satiation and perseverative licking through the use of noncontinuous reinforcement have found a direct relationship between motivation and sucrose concentration (Collier & Siskel, 1959; Guttman, 1953, 1954; Young & Green, 1953; Young & Shuford, 1954). In the present study, with continuous reinforcement, the fact that this relationship is not found in the reinforced BP rate is because of the overlaying effects of physiological satiation and licking behavior.

REFERENCES

CAMPBELL, B. A. Absolute and relative sucrose preferences thresholds for hungry and satiated rats. *J. comp. physiol. Psychol.*, 1958, **51**, 795–799.

COCHRAN, W. G. The comparison of percentages in matched samples. *Biometrika*, 1950, **37**, 256–266.

COLLIER, G., & SISKEL, M. Performance as a joint function of amount of reinforcement and inter-reinforcement interval. *J. exp. Psychol.*, 1959, **57**, 115–120.

GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

GUTTMAN, N. Equal-reinforcement values for sucrose and glucose solutions compared with equal-sweetness values. *J. comp. physiol. Psychol.*, 1954, **47**, 358–361.

MARX, M. H. Resistance to extinction as a function of continuous or intermittent presentation of a training cue. *J. exp. Psychol.*, 1958, **56**, 251–255.

MARX, M. H. Resistance to extinction as a function of degree of reproduction of training conditions. *J. exp. Psychol.*, 1960, **59**, 337–342.

MARX, M. H., TOMBAUGH, J. W., COLE, C., & DOUGHERTY, D. Persistence of nonreinforced responding as a function of the direction of a prior ordered incentive shift. *J. exp. Psychol.*, 1963, **66**, 542–546.

METZGER, R., COTTON, J. W., & LEWIS, D. J. Effect of reinforcement magnitude and order of presentation of different magnitudes on runway behavior. *J. comp. physiol. Psychol.*, 1957, **50**, 184–188.

PUBOLS, B. H., JR. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

SMITH, M., & DUFFY, M. Consumption of sucrose and saccharine by hungry and satiated rats. *J. comp. physiol. Psychol.*, 1957, **50**, 65–69.

YOUNG, P. T., & GREEN, J. T. Quantity of food ingested as a measure of relative acceptability. *J. comp. physiol. Psychol.*, 1953, **43**, 288–294.

YOUNG, P. T., & SHUFORD, E. H. Intensity, duration, and repetition of hedonic processes as related to acquisition of motives. *J. comp. physiol. Psychol.*, 1954, **47**, 298–305.

(Received April 10, 1964)

Journal of Experimental Psychology
1965, Vol. 69, No. 6, 637-638

Beginning with the first issue of Volume 69 and continuing through the remainder of that volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are a portion of the backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Choice Reaction with Variable S-R Mapping: L. H. Shaffer*: Applied Psychology Research Unit, Medical Research Council, 15 Chaucer Road, Cambridge, England.

Competing Responses and the Partial-Reinforcement Effect: Donald F. McCoy, Jr.* and Melvin H. Marx: Department of Psychology, University of Missouri, Columbia, Missouri 65902.

Partial Visual Feedback of Component Motions as a Function of Difficulty of Motor Control: John D. Gould* and Amy Schaffer: IBM No. 73464, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598.

Preference for a Positive Evaluative Response in Concept Learning: Ramon J. Rhine*: Department of Psychology, University of California, Riverside, California.

Learning and Performance as a Function of the Percent Pursuit Component in a Tracking Display: George E. Briggs* and Marty R. Rockway: Office of Research, Ohio State University, 190 Oval Drive, Columbus, Ohio 43210.

Reaction Time and Serial versus Parallel Information Processing: Robert K. Lindsay* and Jane M. Lindsay: 523-B Thompson Avenue, Mountain View, California.

Drive and the Range of Cue Utilization: Donna J. Zaffy and James L. Bruning*: Department of Psychology, Ohio University, Athens, Ohio 45701.

Stimulus versus Response Decisions as Determinants of the Relative Frequency Effect in Disjunctive Reaction-Time Performance: P. John Dillon and Dalbir Bindra*: Department of Psychology, McGill University, Montreal, Canada.

Memory Span as a Function of Variable Presentation Speeds and Stimulus Durations: Michael C. Corballis*: Department of Psychology, McGill University, Montreal, Canada.

Percent Occurrence of Stimulus Members and Meaningfulness as Related to Forward and Backward Recall of Paired Associates: L. R. Goulet* and Robert L. Solso: Institute of Human Learning, University of California, 2241 College Avenue, Berkeley, California 94720.

Effects of Meaningfulness of Structurally Similar CVCs on Stimulus Generalization of Eyelid Closure: David W. Abbott*: Department of Psychology, Stetson University, Deland, Florida 32720.

Prediction and Estimation of a Random Fluctuation: Wayne Lee* and W. R. Garner: Department of Psychology, University of California, Berkeley, California 94720.

Time-Intensity Reciprocity under Various Conditions of Adaptation and Backward Masking: D. Kahneman*: Department of Psychology, Hebrew University of Jerusalem, Jerusalem, Israel.

Effects of Delay of KR and Subject Response Bias on Extinction of a Simple Motor Skill: James A. Dyal*: Department of Psychology, Texas Christian University, Fort Worth, Texas 76129.

Solution of the Two-Stimulus Transposition Problem by Four- and Five-Year-Old Children: Michael D. Zeiler*: Psychological Laboratory, Pendleton Hall, Wellesley College, Wellesley, Massachusetts 02181.

Negative Ionization: An Investigation of Behavioral Effects: J. C. Wofford*: Department of Psychology, University of Southern Mississippi, Southern Station, Box 25, Hattiesburg, Mississippi 39401.

---

* Asterisk indicates author for whom address is supplied.

Conditioning Acceptance or Rejection of Information: B. R. Bugelski* and Michel Hersen: Department of Psychology, Townsend Hall, Administration Road, University of New York, Buffalo, New York 14214.

Effects of Association Value on Perceptual Search: Edward Smith* and Howard Egeth: Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

Some Artifactual Causes of Perceptual Primacy: John J. B. Ayres*: 145 Woodland Avenue, Lexington, Kentucky.

Relationship between Latency and Remoteness in Preference Judgments: Marshall G. Greenberg* and David G. Doren: Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.